INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Functional quantization for pricing derivatives.

Gilles Pagès and Jacques Printems

N° 5392

Novembre 2004

Thème NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5392--FR+ENG

Functional quantization for pricing derivatives.

Gilles Pagès * and Jacques Printems †

Thème NUM — Systèmes numériques
Projet MATHFI

Rapport de recherche n° 5392 — Novembre 2004 — 53 pages

Abstract: We investigate in this paper the numerical performances of quadratic functional quantization and their applications to Finance. We emphasize the rôle played by the so-called product quantizers and the Karhunen-Loève expansion of Gaussian processes. Numerical experiments are carried out on two classical pricing problems: Asian options in a Black-Scholes model and vanilla options in a stochastic volatility Heston model. Pricing based on "crude" functional quantization is very fast and produce accurate deterministic results. When combined with a Romberg log-extrapolation, it always outperforms Monte Carlo simulation for usual accuracy levels.

Key-words: Functional quantization, Product quantizers, Romberg extrapolation, Karhunen-Loève expansion, Brownian motion, SDE, Asian option, stochastic volatility, Heston model.

* Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5. gpa@ccr.jussieu.fr

† Laboratoire d'Analyse et de Mathématiques Appliquées, UMR 8050, Université Paris 12, 61, avenue du Général de Gaulle, F-94010 Créteil & Projet MATHFI INRIA-Rocquencourt. printems@univ-paris12.fr.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Quantification fonctionnelle pour l'évaluation de produits dérivés.

Résumé : Dans cet article, on étudie numériquement les performances de la quantification fonctionnelle quadratique et ses applications à la Finance. On insiste en particulier sur les quantifieurs produits et la série de Karhunen-Loève des processus Gaussiens. Des simulations numériques sont effectuées sur deux problèmes classiques d'évaluation d'options : l'option asiatique dans un cadre Black & Scholes et les options vanilles dans un modèle d'Heston. Le prix obtenu par quantification fonctionnelle brut donne des résultats précis en un minimum de temps. Lorsque cette méthode est combinée avec un extrapolation de Romberg en espace, elle donne de très bons résultats par rapport à Monte Carlo.

Mots-clés : Quantification fonctionnelle, quantifieurs produits, extrapolation de Romberg, série de Karhunen-Loève, mouvement Brownien, EDS, option asiatique, volatilité stochastique, modèle d'Heston.

1 Introduction

This paper is an attempt to investigate the numerical aspects of functional quantization of stochastic processes and their applications to the pricing of derivatives through numerical integration on path-spaces; we will mainly focus on the Brownian motion and the Brownian diffusions viewed as square integrable random vectors defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking their values in the Hilbert space $L^2_T := L^2_{\mathbb{R}}([0,T], dt)$ endowed with the usual norm defined by $|g|_{L^2_T} = (\int_0^T g^2(t)dt)^{1/2}$.

Functional quantization is the natural extension to stochastic processes of the so-called *optimal vector quantization* of random vectors which has been extensively investigated since the late 1940's in Signal processing and Information Theory. Its aim is to provide an optimal spatial discretization of a random vector-valued signal X with distribution $\mathbb{P}_X$ by a random vector taking at most N values $x_1, \ldots, x_N$, called elementary quantizers. Then, instead of transmitting the complete signal $X(\omega)$ itself, one first selects the closest x_i in the quantizer set and transmits its (binary coded) *label* i. After reception, a proxy $\widehat{X}(\omega)$ of $X(\omega)$ is reconstructed using the code book correspondence $i \mapsto x_i$. For a given N, there is (at least) one N-tuple of elementary quantizers which minimizes over $(\mathbb{R}^d)^N$ the quadratic quantization error $\|X - \widehat{X}\|_2$ induced by replacing X by $\widehat{X}$. In d-dimension, this minimal quantization error goes to zero at a $N^{-\frac{1}{d}}$-rate as $N \to +\infty$. Stochastic optimization procedure based on simulation have been devised to compute these optimal quantizers. For an expository of mathematical aspects of quantization in finite dimension we refer to [5] and the references therein. For Signal processing and algorithmic aspects, we refer to [4], [3] and [18].

In the early 1990', optimal quantization has been introduced in Numerical Probability to devise some quadrature integration formulæ with respect to the distribution $\mathbb{P}_X$ on $\mathbb{R}^d$: $\mathbb{E}\,F(X) \approx \mathbb{E}\,F(\widehat{X})$ if N is large enough (even bounded by $[F]_{\text{Lip}}\|X - \widehat{X}\|_2$ if F is Lipschitz continuous) and $\mathbb{E}\,F(\widehat{X})$ is a weighted convex combination of the $F(x_i)$. This approach is efficient in medium dimensions (at least $1 \leq d \leq 4$, see [14], [15] and [18]) especially when many integrals need to be computed against the same distribution $\mathbb{P}_X$: tables of the optimal weighted N-tuples can be computed and kept off-line like as for Gauss points. Later, optimal quantization has been used to design some tree methods in order to solve non-linear problems involving the computation of many conditional expectations: American option pricing, non-linear filtering for stochastic volatility models, portfolio optimization (see [17] for a review of applications to computational Finance).

Recently, the extension of optimal quantization to stochastic processes viewed as random variables taking their values in their path-space has given raise to many theoretical developments (see [11],[12], [2], etc). This *functional quantization* can be seen as a discretization of the path-space of a process, typically the Hilbert space L^2_T. In this paper we aim to develop the numerical aspects of functional quantization and their applications to the pricing of path-dependent derivatives. More precisely, we will focus on additive integral functionals F defined on L^2_T by $\xi \mapsto F(\xi) = \int_0^T f(t,\xi(t))\,dt, \quad \xi \in L^2_T$. The starting point is to search for some "good" computable quantizers (stationnary but possibly not optimal) and then to use the derived quadrature formulæ as a deterministic alternative to Monte Carlo simulation for integrating $\mathbb{E}\,F(X)$ for some process X. In a Gaussian setting, this can be done by using an expansion of X in an appropriate orthonormal basis and, in a non Gaussian setting (like diffusions) by some "quantizing mappings" based on some integral equations (see [13]). Since optimal functional quantization theoretically converges at a rather poor $(\log N)^{-\theta}$-rate for some θ depending on the pathwise regularity of the process X, one has to bet on its performances for "reasonably small" values of N (say $N \leq 10\,000$).

The paper is organized as follows: in Section 2 we provide the reader with some background on quantization of Hilbert spaces and Gaussian processes viewed as L^2_T-valued random vectors. Section 3 is devoted to stationary quantizers and their computational applications (one-dimensional optimal quantizers, etc). Section 4 deals with the weighted quadrature formulæ for the expectations $\mathbb{E}\,F(X)$ of H-valued random vectors X. In Section 5 we investigate a special class of quantizers called scaled product quantizers which will be the key of numerical applications. When X is a Gaussian vector, its Karhunen-Loève (K-L) expansion (*i.e.* its PCA in infinite dimension) plays a crucial rôle. A procedure is described to tabulate the optimal product quantizers. For the Brownian motion these tables are available on the web. In Section 6, a Romberg like method is proposed to speed up numerical integration. In Section 7 we carry out several numerical experiments on two pricing problems: Asian options in a Black-Scholes model and vanilla Calls in a Heston stochastic volatility model. The results are quite promising. In particular, we point out the striking efficiency of a Romberg log-extrapolation which numerically outperforms Monte Carlo simulation in both examples. In Section 8, we outline a kind of FQ-MC method where functional quantization becomes a control variate random variable.

2 Preliminaries on quadratic functional quantization

Let $(H, (.\mid .)_{_H})$ be a separable Hilbert space and $X : (\Omega, \mathcal{A}, \mathbb{P}) \to H$ be a H-valued random vector with distribution $\mathbb{P}_{_X}$ defined on H endowed with its Borel σ-field $\mathcal{B}or(H)$. Typical settings are $H = \mathbb{R}$, $\mathbb{R}^d$ endowed with its canonical Euclidean norm and $L^2_{_T}$ for functional quantization, etc. Quadratic *optimal quantization* consists in studying the best $\|\,.\,\|_{_2}$-approximation of $X \in L^2_{_H}(\Omega, \mathbb{P})$ by H-valued random vectors taking at most N values, including all the induced questions like asymptotic error bounds, rates, construction of nearly optimal quantizers. This framework naturally includes many non-Gaussian processes like diffusions for example (see [13]).

Let $x := (x_1, \dots, x_{_N}) \in H^N$ be a N-quantizer and let $\mathrm{Proj}_x : H \to \{x_1, \dots, x_{_N}\}$ be a projection following the *closest neighbour rule*. It means that the Borel partition $\mathrm{Proj}_x^{-1}(\{x_i\})$, $i = 1, \dots, N$ of H satisfies

$$\mathrm{Proj}_x^{-1}(\{x_i\}) \subset \{\xi \in H \mid |x_i - \xi|_{_H} = \min_{1\le j\le N} |x_j - \xi|_{_H}\},\ 1 \le i \le N.$$

Such a Borel partition is called a *Voronoi tessellation* of H induced by x. The *Voronoi cell* of x is often denoted $C_i(x) := \mathrm{Proj}_x^{-1}(\{x_i\})$. One defines the *Voronoi quantization* of X induced by x by

$$\widehat{X}^x := \mathrm{Proj}_x(X).$$

(the exponent ^{x} will often be dropped or replaced by its size ^{N}) It is the best $L^2(\mathbb{P})$-approximation of X by $\{x_1, \dots, x_{_N}\}$-valued random vectors since, for any random vector $X' : \Omega \to \{x_1, \dots, x_{_N}\}$,

$$\begin{aligned}
\|X - X'\|_{_2}^2 &= \sum_{1\le i\le N} \int_\Omega \mathbf{1}_{C_i(x)}(X(\omega))|X(\omega) - X'(\omega)|^2_{_H}\mathbb{P}(d\omega)\\
&\ge \sum_{1\le i\le N} \int_\Omega \mathbf{1}_{C_i(x)}(X(\omega))\,|X(\omega) - x_i|^2_{_H}\mathbb{P}(d\omega)\\
&= \mathbb{E}(\min_{1\le i\le N} |X - x_i|^2_{_H}) = \|X - \widehat{X}^x\|_{_2}^2
\end{aligned}$$

Note that there are infinitely many Voronoi tessellations which all produce the same quadratic *quantization error* $\|X - \widehat{X}^x\|_{_2}$. In fact the boundaries of the Voronoi cells of any Voronoi tessellation is contained in the same finite union of median hyperplanes $H_{ij} \equiv (x_i - x_j|\,.\,)_{_H} = 0$ $(x_i \ne x_j)$. So, if the distribution $\mathbb{P}_{_X}$ weights no hyperplane, then $\widehat{X}^x$ is $\mathbb{P}$-$a.s.$ uniquely defined.

The second step of the optimization process is to find a N-tuple $x \in H^N$, if any, which minimizes the quantization error over H^N. In fact one checks by the triangular inequality that the function

$$Q_N^X : (x_1, \ldots, x_N) \mapsto \|X - \widehat{X}^x\|_2 = \| \min_{1\le i\le N} |X - x_i|_H \|_2$$

is Lipschitz continuous on H^N. When $N = 1$, $Q_1^2(x) = \mathbb{E}|X - x|_H^2$ is a strictly convex function which reaches its minimum $\mathrm{Var}(|X|_H)$ at $x^* := \mathbb{E}\,X$. Then, one shows by induction on N (see [11] for details), that Q_N^X always reaches a minimum at some optimal N-quantizer $x^* := (x_1^*, \ldots, x_N^*)$. As soon as $|\mathrm{supp}\mathbb{P}_X| \ge N$, any such optimal N-quantizer has pairwise distinct components. The key argument is that the function Q_N^X is weakly lower semi-continuous on H^N. Then, the support of a distribution being σ-compact in the Hilbert space H, it is separable. So, let $(z_n)_{n\ge1}$ denote an everywhere dense sequence in the support of $\mathbb{P}_X$. Then

$$\min_{H^N}(Q_N^X)^2 \le (Q_N^X(z_1, \ldots, z_N))^2 = \int_{\mathrm{supp}(\mathbb{P}_X)} \min_{1\le i\le N} |\xi - z_i|_H^2 \mathbb{P}_X(d\xi) \to 0 \text{ as } N \to \infty$$

by the Lebesgue dominated convergence theorem. Elucidating the rate of convergence of $\min_{H^N} Q_N^X$ toward 0 is a much more demanding problem, even in finite dimension. It has been completely elucidated for non-singular $\mathbb{R}^d$-valued random vectors by the so-called Zador Theorem (see [5]).

Theorem 1 *(Zador, Bucklew & Wise, Graf & Luschgy) Assume that $X \in L^{2+\eta}_{\mathbb{R}^d}(\Omega, \mathbb{P})$ for some $\eta > 0$. Let f denote the density of the absolutely continuous part of $\mathbb{P}_X$ (which can be possibly 0). Then*

$$\min_{(\mathbb{R}^d)^N}(Q_N^X)^2 = \min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_2^2 \sim \frac{J_{2,d}}{N^{2/d}} \left(\int_{\mathbb{R}^d} f^{\frac{d}{d+2}}(\xi) d\xi \right)^{1+2/d} + o\left(\frac{1}{N^{\frac{2}{d}}} \right) \quad \text{as} \quad N \to +\infty.$$

When $f \not\equiv 0$, this yields a sharp rate for the quadratic quantization error since the integral in the right hand side is always finite under the assumption of the theorem. When $f \equiv 0$, this no longer provides a sharp rate, although such sharp rates can be established for some special distributions (self-similar distributions on fractal sets, etc). The true value of $J_{2,d}$ – which corresponds to the uniform distribution over $[0, 1]^d$ – is unknown although one knows that $J_{2,d} = d/(2\pi e) + o(d)$.

In a Hilbert setting no such global result holds, even for Gaussian processes. However, similar sharp rates can be established in some cases when one has a good control on the eigenvalues of the covariance operator of the Gaussian process. The main existing results in that Gaussian setting are brought together in the theorem below.

Theorem 2 *([11] 2002, [12] 2004) Let $H = L^2_T$ and let $(X_t)_{t\in[0,T]}$ be a centered bi-measurable Gaussian process such that $\int_0^T \mathrm{Var}(X_t)dt < +\infty$.*

(a) Let $(e_\ell)_{\ell\geq 1}$ be an orthonormal basis of H. Set $c_\ell^2 := \mathrm{Var}((X|e_\ell)_{L^2_T})$, $n \geq 1$. If there is some real $b > 1$ such that $c_\ell^2 = O(\ell^{-b})$ then,

$$\min_{H^N} Q_N^X = O\left((\log N)^{-\frac{b-1}{2}}\right).$$

(b) Let $(\lambda_\ell)_{\ell\geq 1}$ denote the sequence of eigenvalues of the covariance operator Γ_X of X arranged in an ascending order and let $(e_\ell^X)_{\ell\geq 1}$ be the corresponding orthonormal eigenbasis of H. If there is some real $b > 1$ such that $\lambda_\ell \geq \varepsilon_0\, \ell^{-b}$ for large enough ℓ ($\varepsilon_0 > 0$), then

$$\min_{H^N} Q_N^X \geq \varepsilon_0' (\log N)^{-\frac{b-1}{2}} \qquad \text{for large enough } N \ (\varepsilon_0' > 0).$$

(c) If furthermore, $\lambda_\ell = c_\lambda\, \ell^{-b} + o(\ell^{-b})$ then

$$\min_{H^N} Q_N^X = \frac{c_\lambda^{\frac{1}{2}} b^{\frac{b}{2}}}{2^{\frac{b-1}{2}}(b-1)^{\frac{1}{2}}} (\log N)^{-\frac{b-1}{2}} + o\left((\log N)^{-\frac{b-1}{2}}\right). \tag{2.1}$$

This theorem can be extended by considering the case where c_ℓ^2 and/or λ_ℓ are (upper-bounded by) regularly varying sequences with index $-b$, $b \geq 1$ (and $\sum_\ell c_\ell^2 < +\infty$ when $b = 1$). It turns out that for many (one-parameter) processes, the index b is closely related to the Hölder regularity μ of the application $t \mapsto X_t$ from $[0,T]$ into $L^2(\Omega, \mathbb{P})$: one verifies that $\mu = \frac{b-1}{2}$. For the detailed proofs of the different claims of this theorem in full generality we refer to [11] and [12]. For numerics, the item of interest is (a). Let us emphasize that its proof is constructive and that it *does not rely on optimal quantizers of the process X*. It uses another family of quantizers called *product N-quantizers* which are designed from *optimal quantizers* of the one-dimensional marginals $(X|e_\ell)_{L^2_T}$ of X in the orthonormal basis $(e_\ell)_{\ell\geq 1}$ (see paragraph 5.2 for an outline of the proof). These product quantizers and how to compute them for numerics are in fact the central topic of this paper.

This is the reason why we first provide a short background on numerical methods to obtain optimal quantizers for distributions (Gaussian) on the real line.

3 Stationarity quantizers and first numerical applications

3.1 Smoothness of the distortion function and stationary quantizers

The quantization function Q^X_N is not simply Lipschitz continuous but also differentiable at “most" points of H^N. For convenience let us introduce the *distortion* *i.e.* the square of Q^X_N

$$D^X_N(x) := \mathbb{E} \min_{1\le i\le N} |X - x_i|^2_H, \qquad x\in H^N.$$

Theorem 3 *(a) Let $x := (x_1,\ldots,x_N)\in H^N$ be an N-quantizer satisfying*

$$\forall\, i \neq j, \quad x_i \neq x_j \qquad \textit{and} \qquad \mathbb{P}(X\in \cup_i \partial C_i(x)) = 0 \tag{3.2}$$

($\mathbb{P}_X$-negligible boundary of the Voronoi cells). Then D^X_N is differentiable at x with

$$\frac{\partial D^X_N}{\partial x_i} := 2\,\mathbb{E}(\mathbf{1}_{C_i(x)}(X)(x_i - X)) = 2\int_{C_i(x)} (x_i-\xi)\mathbb{P}_X(d\xi), \quad 1\le i\le N, \tag{3.3}$$

and its differential $\mathrm{D}(D^X_N)$ is continuous. When x^* is an optimal N-quantizer (and $|\mathrm{supp}\mathbb{P}_X|\ge N$), Assumption (3.2) is satisfied (see [5] or [13]) so that

$$\nabla D^X_N(x^*) = 0. \tag{3.4}$$

(b) Assume $H=\mathbb{R}$ and $\mathrm{supp}(\mathbb{P}_X) = \mathrm{closure}((m,M))$ in $\mathbb{R}$, m, $M\in\overline{\mathbb{R}}$. Then, for any (ordered) N-quantizer $x=(x_1,\ldots,x_N)$, $x_1<\ldots<x_N$, its (canonical) Voronoi tessellation is given by

$$C_1(x) = (-\infty, x_{3/2}],\ C_i(x) := (x_{i-1/2}, x_{i+1/2}],\ i=2,\ldots,N-1,\ C_N(x) = (x_{N-1/2}, +\infty) \tag{3.5}$$

where $x_{i-1/2} := \dfrac{x_i + x_{i-1}}{2}$, $i=2,\ldots,N$. If $\mathbb{P}_X$ is absolutely continuous with a continuous p.d.f. f, then D^X_N is twice continuously differentiable at x and its Hessian is given by

$$\mathrm{D}^2(D^X_N)(x) = \left[\frac{\partial^2 D^X_N}{\partial x_i\partial x_j}(x)\right]_{1\le i,j\le N} \tag{3.6}$$

with

$$\frac{\partial^2 D_N^X}{\partial x_i^2}(x)=2\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(u)du - \frac{x_{i+1}-x_i}{2} f(x_{i+\frac{1}{2}})\mathbf{1}_{\{i\le N-1\}} - \frac{x_i - x_{i-1}}{2} f(x_{i-\frac{1}{2}})\mathbf{1}_{\{i\ge 2\}},\ 1\le i\le N,$$

$$\frac{\partial^2 D_N^X}{\partial x_i \partial x_{i-1}}(x)=-\frac{x_i - x_{i-1}}{2} f(x_{i-\frac{1}{2}}),\ 2\le i\le N,\ \frac{\partial^2 D_N^X}{\partial x_i \partial x_{i+1}}(x)=-\frac{x_{i+1}-x_i}{2} f(x_{i+\frac{1}{2}}),\ 1\le i\le N-1,$$

and $\dfrac{\partial^2 D_N^X}{\partial x_i \partial x_j}(x) = 0$ *otherwise.*

(c) *Uniqueness: Assume $H=\mathbb{R}$ and $\mathbb{P}_X$ is absolutely continuous with a* log*-concave p.d.f. f (i.e. $\{f>0\}=(m,M)$, and $\log f$ is concave on (m,M)), then $\{\nabla D_N^X = 0\} = \{x^*\}$.*

The above theorem brings together several classical results about quadratic distortion for which we refer to [5], [15]. It has several consequences on both theoretical and numerical aspects. First it suggests to the following definition of stationary.

Definition 1 *A N-quantizer $x\in H^N$ satisfying (3.2) and the above Equation (3.4) is called a stationary N-quantizer for X. The random vector $\widehat{X}^x$ is called a stationary N-quantization of X.*

If $\mathbb{P}(X\in C_i(x))>0$ for every $i=1,\dots,N$, the equation $\nabla D_N^X(x)=0$ also writes

$$x_i = \frac{\mathbb{E}(\mathbf{1}_{C_i(x)}(X)X)}{\mathbb{P}(X\in C_i(x))} = \mathbb{E}(X \mid \{X\in C_i(x)\}), \qquad i=1,\dots,N. \tag{3.7}$$

Consequently since the σ-fields generated by $\widehat{X}^x$ and $\{\{X\in C_i(x)\},\, i=1,\dots,N\}$ coincide

$$\mathbb{E}(X|\widehat{X}^x) = \widehat{X}^x. \tag{3.8}$$

In particular $\mathbb{E}(X)=\mathbb{E}(\widehat{X}^x)$.

Except for log-concave one-dimensional p.d.f., optimal quantizer(s) are not the only stationary quantizers (see Proposition 4 below about "Karhunen-Loève" product quantizers).

Note that owing to (3.8) the quantization error has then a simpler expression since

$$\mathbb{E}(|X-\widehat{X}^x|_H^2 \mid \widehat{X}^x) \quad = \quad \widehat{X}^x) + |\widehat{X}^x|_H^2$$

$$\mathbb{E}(|X|_H^2 \mid \widehat{X}^x) - |\widehat{X}^x|_H^2 \tag{3.9}$$

$$\text{so that}\quad \|X-\widehat{X}^x\|_2^2 \quad = \quad \mathbb{E}(|X|_H^2) - \mathbb{E}(|\widehat{X}^x|_H^2) = \mathbb{E}(|X|_H^2) - \sum_{1\le i\le N} |x_i|^2 \mathbb{P}(X\in C_i(x)). \tag{3.10}$$

Applying (3.10) to $X - \mathbb{E}\,X$ finally yields

$$\mathbb{E}|X - \widehat{X}^x|_H^2 = \sigma^2_{|X|_H} - \sigma^2_{|\widehat{X}^x|_H} \tag{3.11}$$

where $\sigma_{|Y|_H}$ denotes the standard deviation of $|Y|_H$. We will see further on (Sections 4, 5 and 6) that stationary quantizers are an important class of quantizers for numerics.

Application to processes: When $H = L^2_T$ and X is a bi-measurable process, one also derives from (3.7) that any stationary quantizer has the same regularity as $t \mapsto X_t$ from $[0,T]$ into $L^2(\Omega, \mathcal{A}, \mathbb{P})$ (see [11] and [12] for details).

If, furthermore, X is a Gaussian process, one shows that stationary quantizers lie in the self-reproducing space of X (see [11]). In particular, the components of any stationary quantizer of the Brownian motion all lie in the Cameron-Martin space $H^1 := \{h \in L^2_T \,/\, h(t) = \int_0^t \dot{h}(s)ds,\ \dot{h} \in L^2_T\}$.

3.2 Numerical computation of one-dimensional optimal quantizers

It follows from the former section that, if $\mathbb{P}_X$ is continuous, any optimal (or at least locally optimal) N-quantizer is stationary. This suggests to find them by using a search procedure of the zeros of the gradient ∇D^X_N. In higher dimensions, one usually implement a *stochastic gradient* procedure based on the integral representation of ∇D^X_N (combined with the so-called Lloyd's I procedure, see [18] for details) This has been extensively investigated and experimented in [18], especially for Gaussian vectors. However, as far as one-dimensional log-concave distributions are concerned, extensive numerical experiments carried out in [18] lead to the conclusion that the most efficient procedure is the deterministic Newton-Raphson (NR) algorithm

$$x^{N,(t+1)} = x^{N,(t)} - [\mathrm{D}^2(D^X_N)(x^{N,(t)})]^{-1}\nabla D^X_N(x^{N,(t)}), \quad x^{N,(0)} \in \mathbb{R}^N,$$

where the Hessian $\mathrm{D}^2(D^X_N)(x)$ is given by (3.6). In particular, when dealing with the $\mathcal{N}(0;1)$ distribution on $\mathbb{R}$ (whose p.d.f. is strictly log-concave) the NR-procedure initialized at

$$x^{N,(0)} = \left(-2 + 2\,\frac{2i-1}{2N}\right)_{1\le i\le N}.$$

converges in less than 10 iterates (in the sense that the error reaches the "computer precision") toward the unique optimal N-quantizer x^N. A tabulation of optimal N-quantizers of the

$\mathcal{N}(0;1)$ distribution has been carried out for every $N \in \{1,\dots,400\}$. A file is kept off-line and can be can be downloaded at the URL `www.proba.jussieu.fr/pageperso/pages.html`. It contains

– the (unique) optimal N-quantizer x^N,

– the $\mathbb{P}_\xi$-masses $\mathbb{P}_\xi(C_i(x^N)), i=1,\dots,N$, of its Voronoi cells (*i.e.* the distribution of $\widehat{\xi}^{x^N}$, $\xi \sim \mathcal{N}(0;1)$),

– the induced quadratic quantization error $\|\xi - \widehat{\xi}^{x^N}\|_2$ (using (3.10), given that $\mathrm{Var}(\xi)=1$),

Other quantities of interest like the L^1-quantization error can be computed from these data (closed forms are available, see [18]). (Note that quadratic N-quantizers of the multivariate d-dimensional $\mathcal{N}(0;I_d)$ can be downloaded at the same URL for $d=1,\dots,10$ and various values of N.)

4 Quadrature formulæ for numerical integration

The proposition below illustrates how to use (optimal) quantization for numerical integration of functionals defined on the Hilbert space H: some quadrature formulæ are established with some error bounds. The basic idea is that on the one hand an optimal quantization $\widehat{X}^x$ is close to X *in distribution* and, on the other hand, for every Borel functional $F: H \to \mathbb{R}$ and every $x=(x_1,\dots,x_N) \in H^N$,

$$\mathbb{E}\,F(\widehat{X}^x) = \sum_{1\le i\le N} \mathbb{P}_X(C_i(x))\,F(x_i). \tag{4.12}$$

As soon as one has a numerical access to the N-quantizer x and the distribution $(\mathbb{P}_X(C_i(x)))_{1\le i\le N}$ of the quantization $\widehat{X}^x$, the computation of (4.12) is straightforward. The aim of the proposition below is to establish some error bounds for $\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^x)$ based on L^p-quantization errors $\|X-\widehat{X}^x\|_p$ (with $p=2$ or 4).

Item (a) below is devoted to Lipschitz continuous functionals and item (b) to a second order quadrature formula involving stationary quantizers for smoother functionals. Other quadrature formulæ based on L^p-quantization, $p \neq 2$, can be derived.

Proposition 1 *Let $X \in L^2_H(\Omega,\mathbb{P})$ and let $F: H \to \mathbb{R}$ be a Borel functional defined on H*

(a) First order quadrature formula: *If F is Lipschitz continuous, then*

$$|\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^x)| \le [F]_{\mathrm{Lip}}\|X-\widehat{X}^x\|_2$$

for every N-quantizer $x \in H^N$. In particular, if $(x^N)_{N\geq 1}$ denotes a sequence of quantizers such that $\lim_N \|X - \widehat{X}^{x^N}\|_2 = 0$, then the distribution $\sum_{i=1}^{N} \mathbb{P}_X(C_i(x^N))\delta_{x_i^N}$ of $\widehat{X}^{x^N}$ weakly converges to the distribution $\mathbb{P}_X$ of X as $N \to +\infty$.

(b) SECOND ORDER QUADRATURE FORMULÆ: *Assume that x is a stationary quantizer for X.*

– *Let $\theta : H \to \mathbb{R}_+$ be a nonnegative convex function. If $\theta(X) \in L^2(\mathbb{P})$ and if F is locally Lipschitz with at most θ-growth, i.e. $|F(x) - F(y)| \leq [F]_{\text{Liploc}}|x-y|\,(\theta(x)+\theta(y))$, then $F(X)\in L^1(\mathbb{P})$ and*

$$|\mathbb{E}\, F(X) - \mathbb{E}\, F(\widehat{X}^x)| \leq 2[F]_{\text{Liploc}}\|X - \widehat{X}^x\|_2\|\theta(X)\|_2. \tag{4.13}$$

– *If F is differentiable on H with an α-Hölder differential $\mathrm{D}F$ ($\alpha\in(0,1]$), then*

$$|\mathbb{E}\, F(X) - \mathbb{E}\, F(\widehat{X}^x)| \leq [\mathrm{D}F]_\alpha\|X - \widehat{X}^x\|_2^{1+\alpha}. \tag{4.14}$$

When F is twice differentiable and D^2F is bounded then, one may replace $[\mathrm{D}F]_1 = [\mathrm{D}F]_{\text{Lip}}$ by $\frac{1}{2}\|\mathrm{D}^2F\|_\infty$ in (4.14).

– *If $\mathrm{D}F$ is is locally Lipschitz with at most θ-growth, θ convex, $\theta(X)\in L^4(\mathbb{P})$, then*

$$|\mathbb{E}\, F(X) - \mathbb{E}\, F(\widehat{X}^x)| \leq 3[\mathrm{D}F]_{\text{Liploc}}\|X - \widehat{X}^x\|_4^2\|\theta(X)\|_4. \tag{4.15}$$

(c) AN INEQUALITY FOR CONVEX FUNCTIONALS: *Assume that x is a stationary quantizer. Then for any convex functional $F : H \to \mathbb{R}$*

$$\mathbb{E}\, F(\widehat{X}^x) \leq \mathbb{E}\, F(X). \tag{4.16}$$

The proofs of these quadrature formulæ are postponed to an annex.

Remark: The error bound (4.15) involves $\|X - \widehat{X}^x\|_4$ about which very little is known when x is an optimal (or simply stationary) quadratic quantizer of X: its rate of convergence as N goes to infinity known is not elucidated and numerical computation needs some extra computations. So one often uses a less elegant (and probably less sharp) bound: assume that $\mathrm{D}F$ is is locally Lipschitz with at most θ-growth, θ convex, $\theta(X) \in L^p(\mathbb{P})$ for every $p \geq 1$, then, for every $\varepsilon\in(0,1]$,

$$|\mathbb{E}\, F(X) - \mathbb{E}\, F(\widehat{X}^x)| \leq [\mathrm{D}F]_{\text{Liploc}}\|X - \widehat{X}^x\|_2^{2-\varepsilon}\|X - \widehat{X}^x\|_4^{\varepsilon}(1 + 3\|\theta(X)\|_{\frac{1}{\varepsilon}}). \tag{4.17}$$

Examples: • The typical regular functionals defined on $(L^2_T, |\,.\,|_{L^2_T})$ (most important example for stochastic processes) are the integral functionals F defined by

$$\forall\, \xi \in L^2_T, \qquad F(\xi) = \int_0^T f(t, \xi(t))\, dt$$

where $f : [0, T] \times \mathbb{R} \to \mathbb{R}$ is a Borel function with at most linear growth in x uniformly in t. In particular, F is Lipschitz continuous as soon as $f(t, .)$ is (uniformly in t), convex if $f(t, .)$ is for every t, etc; in particular F is differentiable with an α-Hölder differential as soon as $f(t, .)$ is differentiable for every $t \in [0, T]$ with an α-Hölder partial differential $\frac{\partial f}{\partial x}(t, .)$ (uniformly in t). Then

$$\forall\, \xi \in L^2_T, \qquad \mathrm{D}F(\xi) = \int_0^T \frac{\partial f}{\partial x}(t, \xi(t)) dt.$$

• The functional F defined for every $\xi \in L^2_T$ by

$$F(\xi) := \int_0^T e^{\sigma\xi(t)+\rho t} dt \qquad (\rho \in \mathbb{R})$$

is convex, locally Lipschitz with θ-linear growth, infinitely differentiable. Furthermore, using that $|e^u - e^v| \le |u - v|(e^u + e^v)$ and Schwarz inequality, one derives that

$$[F]_{\mathrm{Liploc}} := \sigma e^{\rho_+ T} \qquad \text{and} \qquad \theta(\xi) = |e^{\sigma\xi}|_{L^2_T}. \tag{4.18}$$

5 Computable rate optimal quantizers for Gaussian processes

In this section, we will focus on (bi-measurable) Gaussian processes X viewed as L^2_T-valued random vectors, although we still consider an abstract Hilbert setting for a while. For convenience we will assume from now on that all random vectors X are centered *i.e.*

$$\mathbb{E}\, X = 0_H.$$

First we will explain why some specific sequences of *product quantizers* with respect to some orthonormal basis $(e_\ell)_{\ell \ge 1}$, although not optimal, produce in some cases the optimal rate of convergence (but not the constant in the sharp rate (2.1)). We will also point out why they cannot be used in general for numerics. As a second step, we will show that when $(e_\ell)_{\ell\geq 1}$ is the eigenbasis of the covariance operator of X (so-called Karhunen-Loève- basis of X), the computation of the distributions of the related quantization $\widehat{X}$ becomes tractable.

5.1 Product quantizers on a Hilbert space

Assume that the Hilbert space $(H, (\,.\,|\,.\,)_H)$ is separable (hence it has a countable orthonormal basis). Let $X \in L^2_H(\Omega, \mathcal{A}, \mathbb{P})$and let $(e_\ell)_{\ell\in L}$ be an orthonormal basis of the Hilbert space H ($L = \{1,\ldots,d\}$ if $\dim H = d$, $L = \mathbb{N}$ otherwise). One may expand X on this basis that is

$$X \stackrel{H}{=} \sum_{\ell\in L} (X|e_\ell)_H\, e_\ell \qquad \mathbb{P}\text{-}a.s.$$

This equality also holds in $L^2_H(\mathbb{P})$. One can normalize this expansion so that

$$X \stackrel{H}{=} \sum_{\ell\in L} c_\ell\, \xi^\ell e_\ell \qquad \mathbb{P}\text{-}a.s. \tag{5.19}$$

with $$c_\ell := \sigma((X|e_\ell)_H) = \sqrt{\mathrm{Var}((X|e_\ell)_H)} \quad \text{and} \quad \xi^\ell := \frac{(X|e_\ell)_H}{c_\ell}\mathbf{1}_{\{c_\ell>0\}},$$

so that the r.v. ξ^ℓ are centered and normalized (when $c_\ell \neq 0$) and $c = (c_\ell) \in \ell^2(L)$.

Definition 2 *Let $N \geq 1$. A N-tuple $x \in H^N$ is called a* product N-quantizer with respect to the orthogonal basis $(e_\ell)_{\ell\in L}$ *if*

$$x := \left(\sum_{\ell\in L} x^{(\ell)}_{i_\ell} e_\ell\right)_{1\leq i_1\leq N_1,\ldots,1\leq i_\ell\leq N_\ell,\ldots}$$

where, for every $\ell \in L$, $x^{(\ell)} := (x^{(\ell)}_1, \ldots, x^{(\ell)}_{N_\ell}) \in \mathbb{R}^{N_\ell}$ is a N_ℓ-quantizer with $N = \prod_{\ell\geq 1} N_\ell$ (hence in infinite dimension, $N_\ell = 1$ and $x^{(\ell)} = 0$ for large enough ℓ). One denotes the component $(x^{(1)}_{i_1}, \ldots, x^{(\ell)}_{i_\ell}, \ldots, 0, 0\ldots)$ of x by $x_{\underline{i}}$ where $\underline{i}$ is the multi-index $(i_1, \ldots, i_\ell, \ldots, 1, 1, \ldots)$.

When there is no ambiguity, one often drops for convenience the reference to the basis $(e_\ell)_{\ell\in L}$ and one denotes the product quantizer by

$$x := \prod_{\ell\in L} x^{(\ell)} = \Big((x^{(1)}_{i_1}, \ldots, x^{(\ell)}_{i_\ell}, \ldots)\Big)_{1\leq i_\ell\leq N_\ell,\, \ell\geq 1}.$$

The "marginals" $x^{(\ell)}$ of such a product quantizer will be used to produce some quantizations $\widehat{\xi^\ell} := \widehat{\xi}^{x^{(\ell)}}$ of the random variables ξ^ℓ appearing in (5.19). In view of (5.19), it is natural to specify a sub-class of product quantizers called *c-scaled product quantizers.*

Definition 3 *Let $x := \prod_{\ell \in L} x^{(\ell)}$ be a product N-quantizer and $c = (c_\ell)_{\ell \in L}$ be a* scaling *vector. The c-scaled product N-quantizer $c \otimes x$ is defined by*

$$c \otimes x := \prod_{\ell \in L} (c_\ell \, x^{(\ell)}) = \left(\sum_{\ell \in L} c_\ell \, x^{(\ell)}_{i_\ell} e_\ell \right)_{1 \le i_1 \le N_1, \ldots, 1 \le i_\ell \le N_\ell, \ldots}.$$

The components of $c \otimes x$ are usually indexed by the multi-index $\underline{i} = (i_1, \ldots, i_\ell, \ldots, 1, 1, \ldots) \in \prod_{\ell \ge 1} \{1, \ldots, N_\ell\}$.

The proposition below describes the simple geometric shape of the Voronoi cells of such a quantizer.

Proposition 2 *Let x be a product N-quantizer and c a scaling vector. Set $\ell_x := \max\{\ell \,/\, N_\ell > 1\} \in \mathbb{N}$.*
(a) Then, the quadratic distortion induced by $c \otimes x$ is

$$D^X_N(c \otimes x) \;=\; \sum_{\ell \ge 1} c_\ell^2 \, D^{\xi^\ell}_{N_\ell}(x^{(\ell)}) \;=\; \mathbb{E}|X|^2 + \sum_{\ell=1}^{\ell_x} c_\ell^2 (D^{\xi^\ell}_{N_\ell}(x^{(\ell)}) - 1). \quad (5.20)$$

(b) Assume that $c_\ell > 0$ for every $\ell \in L$. Then, for every multi-index $\underline{i} \in \prod_{\ell \in L} \{1, \ldots, N_\ell\}$,

$$C_{\underline{i}}(c \otimes x) = \prod_{\ell \in L} (c_\ell C_{i_\ell}(x^{(\ell)})). \quad (5.21)$$

Proof: Both claims follow from the orthonormality of the basis $(e_\ell)_{\ell \in L}$.

$$(a) \qquad \begin{aligned} \mathbb{E} \min_{\underline{i}} |X - (c \otimes x)_{\underline{i}}|^2 &= \mathbb{E} \left(\min_{1 \le i_1 \le N_1, \cdots, 1 \le i_{\ell_x} \le N_{\ell_x}} \left| \sum_{\ell \in L} c_\ell \, \xi^\ell e_\ell - \sum_{\ell=1}^{\ell_x} c_\ell \, x^{(\ell)}_{i_\ell} e_\ell \right|^2 \right) \\ &= \mathbb{E} \left(\min_{1 \le i_1 \le N_1, \cdots, 1 \le i_{\ell_x} \le N_{\ell_x}} \sum_{\ell=1}^{\ell_x} c_\ell^2 \, |\xi^\ell - x^{(\ell)}_{i_\ell}|^2 + \sum_{\ell \ge \ell_x + 1} c_\ell^2 \, (\xi^\ell)^2 \right) \\ &= \sum_{\ell=1}^{\ell_x} c_\ell^2 \, \mathbb{E} \left(\min_{1 \le i_{\ell_x} \le N_{\ell_x}} |\xi^\ell - x^{(\ell)}_{i_\ell}|^2 \right) + \sum_{\ell \ge \ell_x + 1} c_\ell^2. \end{aligned}$$

The first equality follows from the fact that, for every $\ell > \ell_x$, $x^{(\ell)} = \mathbb{E}(\xi^\ell) = 0$ so that $D_1^{\xi^\ell}(0) = \mathrm{Var}(\xi^\ell) = \mathbf{1}_{\{c_\ell > 0\}}$.
(b) One may assume without loss of generality that, for every $\ell \in L$, the components of $x^{(\ell)}$ are in an ascending order *i.e.* $i \mapsto x_i^{(\ell)}$ is nondecreasing. Let $\underline{i} := (i_1, \ldots, i_{\ell_x}, 1, \ldots)$ and $\underline{j} := (j_1, \ldots, j_{\ell_x}, 1, \ldots)$. Then, if $\zeta = \sum_\ell \zeta_\ell e_\ell \in H$, $|\zeta - (c\otimes x)_{\underline{i}}|^2 < |\zeta - (c\otimes x)_{\underline{j}}|^2$ iff

$$\sum_{\ell=1}^{\ell_x} c_\ell^2 \left(x_{i_\ell}^{(\ell)} - x_{j_\ell}^{(\ell)}\right)\left(\frac{\zeta_\ell}{c_\ell} - \frac{x_{i_\ell}^{(\ell)} + x_{j_\ell}^{(\ell)}}{2}\right) < 0.$$

Then, for every fixed ℓ, setting $j_\ell = i_{\ell \pm 1}$ and $j_{\ell'} = i_{\ell'}$ if $\ell' \neq \ell$ implies that

$$\widetilde{x}_{i_\ell}^{(\ell)} < \frac{\zeta_\ell}{c_\ell} < \widetilde{x}_{i_{\ell+1}}^{(\ell)} \quad i.e. \quad \frac{\zeta_\ell}{c_\ell} \in C_{i_\ell}(x^{(\ell)}).$$

One checks that this condition is sufficient. $\diamond$

Corollary 1 (a) *It is possible to rearrange the orthonormal basis (e_ℓ) so that sequence $(c_\ell)_{\ell\geq 1}$ is non-increasing. Assuming this has been done, one has*

$$\min_{H^N} D_N^X \leq \min\left\{\sum_{\ell=1}^m c_\ell^2 \min_{\mathbb{R}^{N_\ell}} D_{N_\ell}^{\xi^\ell} + \sum_{\ell \geq m+1} c_\ell^2,\ N_1 \times \cdots \times N_m \leq N,\ N_1, \ldots, N_m \geq 2,\ m \geq 1\right\}. \tag{5.22}$$

(b) *Let $x = \prod_{\ell \in L} x^{(\ell)} \in H^N$, be a product N-quantizer, let $c = (c_\ell) \in \ell^2(L)$ with $c_\ell > 0$, $\ell \in L$. Let $\widehat{\xi}^\ell := \widehat{\xi}^{x^{(\ell)}}$ be the (Voronoi) quantization of ξ^ℓ induced by $x^{(\ell)}$. The $c\otimes x$-quantization of X is given by*

$$\widehat{X}^{c\otimes x} = \sum_{\ell \geq 1} c_\ell \widehat{\xi}^\ell e_\ell = \sum_{\underline{i}} (c\otimes x)_{\underline{i}} \mathbf{1}_{C_{\underline{i}}(c\otimes x)}(X), \quad \text{with} \quad C_{\underline{i}}(c\otimes x) = \prod_{\ell \geq 1}\{\xi^\ell \in C_{i_\ell}(x^{(\ell)})\}. \tag{5.23}$$

Proof: (a) The rearrangement is possible since $c \in \ell^2(L)$. The claim follows from Proposition 2(a).
(b) It follows from (5.21) that

$$X \in C_{\underline{i}}(x) \quad \text{iff} \quad c_\ell \xi^\ell \in c_\ell C_{i_\ell}(x^{(\ell)}),\ \ell \geq 1 \quad \text{iff} \quad \xi^\ell \in C_{i_\ell}(x^{(\ell)}),\ \ell \geq 1. \quad \diamond$$

Remark. The choice of rearranging the coefficients c_ℓ in a descending order is natural. Moreover, it is established in [11] (Theorem 3.2 and the remarks below) that when $(e_\ell)_{\ell \in L}$ is the Karhunen-Loève basis (see Section 5.3 below), this choice is the optimal one.

5.2 Product quantizers of a Gaussian process

The estimate in (a) is at the origin of the asymptotic upper-bounds of the quadratic quantization error for Gaussian processes first established in [11]. We will briefly recall the approach in order to explain first how it may produce the right asymptotic rate of decay of the quadratic quantization error and second why it cannot be used in such a generality for numerical purpose. Indeed, if $H = L^2_T$ and X is a bi-measurable 0-centered Gaussian process such that

$$\int_0^T \mathbb{E}\, X_t^2\, dt < +\infty, \tag{5.24}$$

then X can be seen as a L^2_T-valued random vector. Furthermore, X being a Gaussian process, the sequence of random variables $(\xi^\ell)_{\ell\geq 1}$ is a 0-centered Gaussian sequence of normal random variables. In particular, for every $\ell \in L$ there exists $x^{(\ell)} \in \mathbb{R}^{N_\ell}$ such that

$$D^{\xi}_{N_\ell}(x^{(\ell)}) = \min_{\mathbb{R}^{N_\ell}} D^{\xi^\ell}_{N_\ell} = \min_{\mathbb{R}^{N_\ell}} D^{\xi}_{N_\ell}(x^{(\ell)}), \qquad \xi \sim \mathcal{N}(0;1).$$

Consequently, there exists a real constant $K > 0$ given by Zador's Theorem such that

$$D^{\xi}_{N_\ell}(x^{(\ell)}) = \min_{\mathbb{R}^{N_\ell}} D^{\xi^\ell}_{N_\ell} \leq K\, N_\ell^{-2}, \qquad \ell \in L. \tag{5.25}$$

Then, if one is interested in the rate of convergence of $\min_{(L^2_T)^N} D^X_N$ as $N \to \infty$, one may replace the optimization problem in the right hand side of (5.22) by the *optimal size allocation problem* (still assuming (c_ℓ) is non-increasing), namely

$$\min\left\{\sum_{\ell=1}^{m} \frac{c_\ell^2}{N_\ell^2} + \sum_{\ell\geq m+1} c_\ell^2,\ N_1\times\cdots\times N_m \leq N,\ N_1,\ldots,N_m \geq 2,\ m\geq 1\right\}. \tag{5.26}$$

Set $m_N := \max\left\{ m\geq 1 \ \text{ s.t. }\ N^{\frac{1}{m}} c_m \left(\prod_{\ell=1}^{m} c_\ell\right)^{-\frac{1}{m}} \geq 1\right\}$, $[u] := \max\{k\in\mathbb{N} \,/\, k \leq u\}$ and, then,

$$N_\ell := \left[N^{\frac{1}{m_N}} c_\ell \left(\prod_{k=1}^{m_N} c_k\right)^{-\frac{1}{m_N}}\right],\ \ell = 1,\ldots,m_N, \qquad N_\ell = 1,\ \ell \geq m_N + 1, \quad N_\ell := 1. \tag{5.27}$$

The resulting sequence $(N_\ell)_{\ell\geq 1}$ of quantizer sizes (whose product is at most N for every N) is asymptotically optimal for (5.26). If $c_\ell = O(\ell^{-b})$ one derives that $m_{_N} = \frac{2}{b}\log N + o(\log N)$ and that the value function in (5.26) goes to 0 at a $O((\log N)^{-\frac{b-1}{2}})$-rate (see [11], [16], [12] for details). This yields the asymptotic rate of decay for $\|X - \widehat{X}^{c\otimes x^N}\|_2$ as $N\to\infty$ where *x^N is a product quantizer whose "marginals" $x^{(\ell)}$ are N_ℓ-optimal with N_ℓ given by (5.27).* This completes the proof of Theorem 2 (a). In fact, it even yields the slightly more accurate result (see [12], Theorem 2.2 for details). Let

$$\mathcal{O}_{pq}(N) := \left\{\prod_{\ell\geq 1} x^{(\ell)},\ x^{(\ell)} \text{ optimal } N_\ell\text{-quantizer of the } \mathcal{N}(0;1)\text{-distribution}, \ell\geq 1,\ \prod_{\ell\geq 1} N_\ell \leq N\right\}. \tag{5.28}$$

where the subscript pq means *product quantizer.*

Proposition 3 *Let X be a Gaussian process. Assume that $c_n^2 \leq c^*\, n^{-b}$, $b>1$. Then*

$$\min\left\{\|X - \widehat{X}^{c\otimes x}\|_2, x\in \mathcal{O}_{pq}(N)\right\} \leq \left(c^*\left(\frac{b}{2}\right)^{b-1}\left(\frac{1}{b-1} + 4C_{\mathcal{N}(0;1)}\right)\right)^{1/2} \frac{1}{(\log N)^{\frac{b-1}{2}}} \tag{5.29}$$

where

$$C_{\mathcal{N}(0;1)} := \sup_{N\geq 1}\left(N^2 \min_{y\in\mathbb{R}^N} D_{_N}^{\mathcal{N}(0;1)}(y)\right).$$

A conjecture (see [12]) is that in fact $C_{\mathcal{N}(0;1)} = \lim_N \left(N^2 \min_{x\in\mathbb{R}^N} D_{_N}^{\mathcal{N}(0;1)}(x)\right) = \dfrac{\pi}{2}\sqrt{3}$. (The second equality follows from Zador's Theorem.) Numerical experiments tend to confirm this conjecture: the inequality $\max_{n\leq N}(n^2 \min_{x\in\mathbb{R}^n} D_n^{\mathcal{N}(0;1)}(x)) \leq \dfrac{\pi}{2}\sqrt{3}$ is satisfied at least up to $N = 10\,000$.

Consequently, if the rate of convergence of (c_ℓ) as $\ell\to\infty$ is known ($e.g.$ in a power scale) for a given orthonormal basis (e_ℓ) of $L^2_{_T}$, one derives a rate of decay of the quantization error $\min_{x\in\mathcal{O}_{pq}(N)}\|X - \widehat{X}^{c\otimes x^N}\|_2$ as $N\to\infty$ with, sometimes, some numerical estimates for finite N.

Thus for the fractional Brownian motion on the unit interval $W^\alpha = (W_t^\alpha)_{t\in[0,1]}$ with Hurst constant α, one may consider the Haar basis defined as the restriction on $[0,1]$ of the functions

$$e_0 := \mathbf{1},\ e_1 := \mathbf{1}_{[0,1/2)} - \mathbf{1}_{[1/2,1]},\ e_{2^k+\ell} := 2^{k/2} e_1(2^k t - \ell),\ \ell = 0,\ldots,2^k-1,\ k\in\mathbb{N}.$$

Using the self-similarity property one derives that

$$c_{2^k+\ell} = \frac{c_1}{2^{k(\alpha+1/2)}}, \quad \ell = 0, \dots, 2^k - 1, \; k \geq 1.$$

so that an appropriate choice of m, N_1,..., N_m (as functions of N) yields

$$\min_{\chi \in (L^2_T)^N} \|W^\alpha - \widehat{W^\alpha}^{\chi}\|_2 \leq \min_{x \in \mathcal{O}_{pq}(N)} \|W^\alpha - \widehat{W^\alpha}^{c\otimes x}\|_2 \leq \frac{K'_\alpha}{(\log N)^\alpha}.$$

Many other quantization rates can be obtained that way (*e.g.* for stationary processes by considering the trigonometric basis $(e^{iu\cdot})_{u\in\mathbb{R}}$, for multi-parameter processes, etc, see [11] and [12]).

This general approach has a major drawback for numerics: although the Voronoi cells $C_{\underline{i}}(c \otimes x)$ do have a simple geometric shape given by Equation (5.23) and even if $(e_\ell)_{\ell\geq 1}$ and the coefficients $(c_\ell)_{\ell\geq 1}$ are known, one problem remains for numerical purposes: the *distribution* of $\widehat{X}^{c\otimes x}$, *i.e.*

$$\mathbb{P}(\widehat{X}^{c\otimes x} = (c \otimes x_{\underline{i}})) = \mathbb{P}\left(\bigcap_{\ell\in L} \{\xi^\ell \in C_{i_\ell}(x^{(\ell)})\}\right), \qquad \underline{i} \in \prod_{\ell\geq 1}\{1, \dots, N_\ell\}, \tag{5.30}$$

cannot be computed simply in practice. The sequence $(\xi^\ell)_{\ell\in L}$ is Gaussian, every ξ^ℓ is centered with variance $\mathbf{1}_{\{c_\ell>0\}}$ but its distribution is characterized by its covariance structure given by

$$\mathrm{Cov}(\xi^\ell, \xi^{\ell'}) = \frac{\int_{[0,T]^2} e_\ell(s) e_{\ell'}(t) \mathbb{E}(X_s X_t)\, ds\, dt}{c_\ell\, c_{\ell'}} \mathbf{1}_{\{c_\ell, c_{\ell'}>0\}}, \qquad \ell,\, \ell' \in L. \tag{5.31}$$

For a generic orthonormal basis $(e_\ell)_{\ell\in L}$, no closed form is available for such quantities, even if an explicit expression for the covariance function $(s,t) \mapsto \mathbb{E}(X_s X_t)$ is available. For the same reason a Monte Carlo simulation based on Expansion (5.19) cannot be implemented at a reasonable cost.

However, there is a specific orthonormal basis of L^2_T closely related to the process X for which this problem can be overcome.

5.3 Product quantizers of the Karhunen-Loève expansion of a Gaussian process

As soon as a Gaussian process satisfies (5.24), there is a special orthonormal basis associated to a bi-measurable square integrable Gaussian process $(X_t)_{t\in[0,T]}$: the eigenbasis associated to its covariance operator by

$$\forall\, f \in L^2_T, \qquad \Gamma_X(f) := \left(t \mapsto \int_0^T f(s)\,\mathbb{E}(X_t X_s)ds\right).$$

The operator Γ_X is a non-negative self-adjoint compact operator which can be diagonalized in an orthonormal basis $(e^X_\ell)_{\ell\geq 1}$ of L^2_T:

$$\Gamma_X(e^X_\ell) = \lambda_\ell e^X_\ell, \qquad \ell \geq 1,$$

where the eigenvalues make up a nonincreasing sequence $(\lambda_\ell)_{\ell\geq 1}$ of nonnegative real numbers. Without loss of generality one may assume that

$$\forall\, \ell \in L, \qquad \lambda_\ell > 0 \tag{5.32}$$

since otherwise $\mathrm{supp}X \neq H$. Then, this eigenbasis is unique. In case X is in fact a finite dimensional Gaussian vector, then, most of what follows remains true by setting $d := \min\{\ell \geq 1 \,/\, \lambda_\ell > 0\}$ and considering $L := \{1,\ldots,d\}$ instead of $\{1,\ldots,\ell,\ldots\}$ as an index set.

Consequently

$$\forall\, f,\, g \in L^2_T,\ \mathrm{Cov}\left((f|X)_{L^2_T}, (g|X)_{L^2_T}\right) = \int_{[0,T]^2} f(t)g(s)\mathbb{E}(X_tX_s)ds\,dt = (f|\Gamma_X(g))_{L^2_T} \tag{5.33}$$

so that
$$c^2_\ell := \mathbb{E}((X|e^X_\ell)^2_{L^2_T}) = \left(e^X_\ell|\Gamma_X(e^X_\ell)\right)_{L^2_T} = \lambda_\ell \tag{5.34}$$

and
$$\mathrm{Cov}(\xi^\ell, \xi^{\ell'}) = \frac{(e^X_\ell|\Gamma_X(e^X_{\ell'}))_{L^2_T}}{c_\ell\, c_{\ell'}} = \delta_{\ell,\ell'}. \tag{5.35}$$

Then: – the sequence $(\xi^\ell)_{\ell\geq 1}$ is now *i.i.d.* with standard normal distribution (white Gaussian noise),

– the expansion (5.19) related to $(e^X_\ell)_{\ell\in L}$ now reads

$$X_t(\omega) \stackrel{L^2_T}{=} \sum_{\ell\geq 1} \sqrt{\lambda_\ell}\,\xi^\ell(\omega)e^X_\ell(t) \qquad \mathbb{P}(d\omega)\text{-}a.s. \qquad (\text{and in } L^2([0,T]\times\Omega, d\,\mathbb{P}\otimes dt)) \tag{5.36}$$

This expansion is known as the *Karhunen-Loève* (K-L) expansion of X and the basis (e^X_ℓ) as the *Karhunen-Loève* basis. It is the PCA of the L^2_T-valued random variable X: for every

$k \geq 1$, $\mathrm{Var}(|\mathrm{Proj}^{\perp}_{<e_1^X,\ldots,e_k^X>}(X)|_{L^2_T}) = \lambda_1 + \cdots + \lambda_k$ is maximum among all orthonormal basis of L^2_T. Another feature of this expansion is the simultaneous *orthonormality of the basis* $(e_\ell^X)_{\ell\in L}$ *and the independence of the* $\xi^\ell, \ell \in L$. This implies by a straightforward martingale argument that Equality (5.36) is also true $\mathbb{P}(d\omega)$-*a.s.* at dt-almost every time $t \in [0,T]$. Moreover, it makes the K-L expansion quite appropriate for Monte Carlo simulation of X (once truncated).

As concerns functional quantization, the $\sqrt{\lambda}$-scaled product quantizer $\sqrt{\lambda}\otimes x$ related to the K-L basis and the resulting quantization $\widehat{X}_t^{\sqrt{\lambda}\otimes x}$ of X read

$$\sqrt{\lambda}\otimes x = \sum_{\ell\in L}\sqrt{\lambda_\ell}\, x^{(\ell)}_{i_\ell} e_\ell^X \quad \text{and} \quad \widehat{X}_t^{\sqrt{\lambda}\otimes x} = \sum_{\ell\geq 1}\sqrt{\lambda_\ell}\,\widehat{\xi}^\ell e_\ell^X(t), \qquad \widehat{\xi}^\ell := \widehat{\xi}^{x^{(\ell)}},\ \ell \geq 1, \tag{5.37}$$

respectively. But the key point is that now that *the distribution (5.30) of* $\widehat{X}^{\sqrt{\lambda}\otimes x}$ *is computable* since

$$\forall\, \underline{i} \in \prod_{\ell\geq 1}\{1,\ldots,N_\ell\}, \qquad \mathbb{P}(\widehat{X}^{\sqrt{\lambda}\otimes x} = (\sqrt{\lambda}\otimes x)_{\underline{i}}) = \prod_{\ell\geq 1}\mathbb{P}(\xi \in C_{i_\ell}(x^{(\ell)})), \qquad \xi \sim \mathcal{N}(0;1). \tag{5.38}$$

(Keep in mind that $(\mathbb{P}(\xi\in C_i(x^{(\ell)})))_{i=1,\ldots,N_\ell}$ is simply the distribution of $\widehat{\xi}^{x^{(\ell)}}$, $\xi\sim\mathcal{N}(0;1)$). These quantities which are easy to compute using (3.5) are already kept off line up to $N_\ell = 1\,000$ (available at the above URL, see paragraph 3.2). For our purpose here, no values for N_ℓ up to 30 are sufficient.

In some sense we moved from the numerical tractability of the distribution of the sequence $(\xi^\ell)_{\ell\in L}$ in a given orthonormal basis to the determination of the K-L basis of a Gaussian process. This question cannot be solved numerically in full generality either; however for many important processes as illustrated below this basis can be made explicit. Before passing to these fundamental examples, let us mention two specific features of the K-L expansion for quantization. First, the lower bounds for the quantization rates are based on some entropy estimates derived from the rate of convergence of the eigenvalues λ_ℓ to 0 (see [11] and [12]). The second one is that, among all orthonormal basis of L^2_T the K-L one preserves the stationarity in the following sense.

Proposition 4 *Let $x^{(\ell)}, \ell\in L$, denote a family of stationary N_ℓ-quantizers of the normal distribution such that $N_\ell = 1$ for every large enough ℓ (i.e. $x^{(\ell)} = 0$). Then the $\sqrt{\lambda}$-scaled product quantizer $\sqrt{\lambda}\otimes x$ is a stationary quantizer for X.*

Proof (See also [5], Lemma 4.8). The ξ^ℓ being independent, the $\widehat{\xi}^\ell$ are independent too. Furthermore, it is obvious from (5.37) and the identity $\widehat{\xi}^\ell = (\widehat{X}^{\sqrt{\lambda}\otimes x}|e_\ell^X)_{L^2_T}/\sqrt{\lambda_\ell}$, $\ell\in L$, that $\sigma(\widehat{X}^{\sqrt{\lambda}\otimes x}) = \sigma(\widehat{\xi}^\ell,\, \ell\in L)$. Consequently

$$\begin{aligned}
\mathbb{E}(X|\widehat{X}^{\sqrt{\lambda}\otimes x}) &= \sum_\ell \sqrt{\lambda_\ell}\,\mathbb{E}(\xi^\ell|\,\widehat{\xi}^\ell,\,\ell\in L)\,e_\ell^X = \sum_\ell \sqrt{\lambda_\ell}\,\mathbb{E}(\xi^\ell|\,\widehat{\xi}^\ell,\,\widehat{\xi}^{\ell'},\,\ell'\in L,\,\ell'\neq\ell)\,e_\ell^X\\
&= \sum_\ell \sqrt{\lambda_\ell}\,\mathbb{E}(\xi^\ell|\,\widehat{\xi}^\ell)e_\ell^X = \sum_\ell \sqrt{\lambda_\ell}\,\widehat{\xi}^\ell e_\ell^X \qquad \text{by the stationarity of } x^{(\ell)} \text{ for } \xi^\ell,\\
&= \widehat{X}^{\sqrt{\lambda}\otimes x}. \qquad \diamond
\end{aligned}$$

Basic examples of interest: – *The Brownian motion on* $[0,T]$. The Karhunen-Loève eigenbasis and its eigenvalues admit a closed form given by

$$e_\ell^W(t) := \sqrt{\frac{2}{T}}\sin\left(\pi(\ell-1/2)\frac{t}{T}\right), \qquad \lambda_\ell := \left(\frac{T}{\pi(\ell-1/2)}\right)^2, \qquad \ell\geq 1. \tag{5.39}$$

– *The Brownian bridge.* The Karhunen-Loève eigenbasis and its eigenvalues are given by

$$e_\ell^X(t) := \sqrt{\frac{2}{T}}\sin\left(\pi\ell\frac{t}{T}\right), \qquad \lambda_\ell := \left(\frac{T}{\pi\,\ell}\right)^2 \qquad \ell\geq 1. \tag{5.40}$$

5.4 Numerical optimization when the K-L expansion is explicit

5.4.1 The “blind" optimization procedure

Assume that *closed forms* are available for both the eigensystem $(\lambda_\ell, e_\ell^X)_{\ell\geq 1}$ of a Gaussian process X as it is the case for both Brownian motion and Brownian bridge as emphasized above.

Then, we are in the position to solve numerically the optimization problem appearing at the right hand side of (5.22) for every $N\in\{1,\ldots,N_{\max}\}$ (so far, we reached $N_{\max} := 11\,519$) and then to compute the “companion parameters" of the optimal or *record* scaled product quantizer $\sqrt{\lambda}\otimes x^N_{\rm rec}$ (distribution of $\widehat{X}^{\sqrt{\lambda}\otimes x^N_{\rm rec}}$, quantization error $Q^X_N(\sqrt{\lambda}\otimes x^N_{\rm rec})$). The reason for implementing such a *blind* optimization procedure is that the values for N_ℓ and m_N given in (5.27) (once set $c_\ell = \sqrt{\lambda_\ell}$) are only asymptotically optimal. For numerical applications, we are interested in this optimization for small values of N. This “blind" optimization procedure is carried out in two steps.

Phase 1(Optimization phase at fixed N): Producing for every N a $\sqrt{\lambda}$-scaled product N-quantizer $\sqrt{\lambda}\otimes x_{\rm opt}$ which is *optimal* in the sense that it induces the lowest quadratic quantization error among all scaled product quantizers *of size exactly* N.

The components $x^{(\ell)}$ of the product quantizer x are optimal quantizers of the normal distribution $\mathcal{N}(0;1)$. So, to compute $\sqrt{\lambda}\otimes x_{\rm opt}$, the distribution of $\widehat{X}^{\sqrt{\lambda}\otimes x_{\rm opt}}$ and the induced quadratic quantization error $Q_N^X(\sqrt{\lambda}\otimes x_{\rm opt})$, we simply need to use the "library" storing the optimal N_ℓ-quantizers $x^{(*,N_\ell)}$, their distributions $(\mathbb{P}(\xi\in C_i(x^{(*,N_\ell)})))_{1\le i\le N_\ell}$ and the quadratic quantization errors $Q_{N_\ell}^{\mathcal{N}(0;1)}(x^{(*,N_\ell)})$. The computation of the quadratic quantization error is based on formula (3.10) for distortion. In practice, $N_\ell\le 100$ is enough for values of N as high as 10^6 since decompositions involving not enough factors will clearly be far from optimality (this heuristic rule is based on the theoretical formula (5.27) below for the optimal values of N_ℓ as N goes to infinity).

Phase 2 (Record Selection phase): Storing for every $N\in\{1,\ldots,N_{\max}\}$,

– the size $N_{\rm rec}:=N_{\rm rec}(N)\in\{1,\ldots,N\}$ which produces the lowest quadratic quantization error,

– the optimal decomposition $N_{\rm rec}=N_1^{\rm rec}\times\cdots\times N_\ell^{\rm rec}\times\cdots\times N_{\ell_{\rm rec}}^{\rm rec}$, (with $N_\ell^{\rm rec}\ge 2$),

– the product quantizer $x_{\rm rec}^N:=x_{\rm opt}^{N_{\rm rec}}$ ($x_{\rm rec}^N=\prod_{\ell\ge1}x^{(\ell)}$, $x^{(\ell)}$ optimal $N_\ell^{\rm rec}$-quantizer of $\mathcal{N}(0;1)$, $\ell\ge1$) so that the $\sqrt{\lambda}$-scaled product $N_{\rm rec}$-quantizer $\sqrt{\lambda}\otimes x_{\rm rec}$ solves the optimization problem at level N.

– the distribution of $\widehat{X}^{\sqrt{\lambda}\otimes x_{\rm rec}^N}$ (using formula (5.38)),

– the corresponding quantization error (and distortion) using (5.20) (setting $c=\lambda$).

Table 1 below provides the first three quantities for some values of N, namely $N=1$, 10, 100, 1 000, 10 000 and 11 519 (the full record table, the record quantizer list including the distributions are available at the same URL up to 11 519). Figures 1, 2, 3 show the scaled product quantizers of the Brownian motion on $[0,1]$ for $N=10,48$ and for the "record value" of $N=100$ that is $N_{\rm rec}=96$.

N	$N_{\rm rec}$	Quant. Error	Opti. Decomp.
1	1	0.7071	1
10	10	0.3138	5 – 2
100	96	0.2264	12 – 4 – 2
1 000	966	0.1881	23 – 7 – 3 – 2
10 000	9 984	0.1626	26 – 8 – 4 – 3 – 2 – 2
11 519	11 232	0.1617	26 – 9 – 4 – 3 – 2 – 2

Table 1 *Brownian motion: Some typical "record" values for numerical implementations*

Fig.2 shows the graphs of both $N \mapsto D^W_{N_{\rm rec}}(\sqrt{\lambda}\otimes x^N_{\rm rec})$ and $N \mapsto D^W_N(\sqrt{\lambda}\otimes x_{\rm opt})$ for $N \in \{1,\ldots,1\,000\}$. Fig.3 depicts $\log(N) \mapsto (D^W_{N_{\rm rec}}(\sqrt{\lambda}\otimes x^N_{\rm rec}))^{-1}$ which emphasizes $\log N$ behaviour of the distortion. The coefficients obtained by linear regression yield

$$\frac{1}{D^W_{N_{\rm rec}}(\sqrt{\lambda}\otimes x^N_{\rm rec})} \approx 4\log N + 2 \quad i.e. \quad D^W_{N_{\rm rec}}(\sqrt{\lambda}\otimes x^N_{\rm rec}) \approx \frac{0.25}{\log N + 0.5},\; 1 \le N \le 10\,000.$$

The lower and upper bounds provided by (2.1) and (5.29) respectively are on $[0,1]$,

$$\frac{1}{\pi^2}\frac{2^2}{2^{2-1}(2-1)} = \frac{2}{\pi^2} \approx 0.2026 < \;\mathbf{0.25} < 1.2040 \approx \frac{1}{\pi^2}(1+2\pi\sqrt{3}). \tag{5.41}$$

This points out that optimal scalar quantizers are nearly globally optimal for Brownian Motion (at least within this range of values for N). For higher values of N one may give up the blind optimization procedure and rely on the asymptotic optimal sizes given by (5.27). Proposition 3 that is

One may even speed up this approach by using the asymptotic estimate $m_{_N} \sim \log(N)$ for large values of N (since $b = 2$ for the Brownian Motion). One may note that in fact

$$N_\ell \approx \sqrt{\frac{\lambda_\ell}{\lambda_{\ell_N}}},\; \ell = 1,\ldots,m_{_N}.$$

In practice, the product $N' = N_1 \times \cdots \times N_{\ell_N}$ can be significantly lower than N, especially when N is not too large, owing to the truncation effect. So a more efficient choice is to consider the upper truncation instead of the regular integral value in (5.27) although this time N' can be ...significantly larger than N. Furthermore, note that N' is *a priori* not a record integer $N_{\rm rec}$ and that this very decomposition can be sub-optimal for N'. Table 2 below gives some examples of such decompositions corresponding to values of N beyond the shortcoming of the optimization procedure at $N_{\max} = 11\,519$: thus $N = 13\,500$ in Table 2 provides worse results than $N = 11\,232$ in Table 1.

N	Quant. Error	Decomp
13 500	0.16217	25 − 9 − 5 − 3
40 500	0.15362	25 − 9 − 5 − 4 − 3 − 3
104 400	0.14811	29 − 10 − 6 − 5 − 4 − 3
313 200	0.14140	29 − 10 − 6 − 5 − 4 − 3 − 3

Table 2. *Brownian motion: some decompositions for higher values of N*

5.4.2 Application to computable rate optimal quantizers for the antiderivative of the Brownian motion

We will illustrate in this short paragraph how rate optimal product quantizers of the Brownian motion can produce some (non Voronoi) rate optimal quantizers of its antiderivative. This process is involved in the control variate variable of the Asian Call (see paragraph 7.1).

First note that one can integrate a Karhunen-Loève expansion of the Brownian motion. In fact, $h \mapsto \int_0^\cdot h(s)ds$ being a Lipschitz continuous function from L^2_T into $(C([0,T]), \|\,.\,\|_{\sup})$, one has, in $L^2_{(C([0,T]),\|\cdot\|_{\sup})}(\mathbb{P})$ (and $\mathbb{P}$-$a.s.$ in L^2_T):

$$\int_0^t W_s ds \overset{L^2_T}{=} \sum_{\ell\ge1} \lambda_\ell\,\xi^\ell \sqrt{\frac{2}{T}}\left(1-\cos\left(\frac{t}{\sqrt{\lambda_\ell}}\right)\right) \quad \text{with} \quad \lambda_\ell := \left(\frac{T}{\pi(\ell-1/2)}\right)^2, \ \ell\ge1, \tag{5.42}$$

$$= 2\sqrt{\frac{2}{T}}\sum_{\ell\ge1}\lambda_\ell\,\xi^\ell \sin^2\left(\frac{t}{2\sqrt{\lambda_\ell}}\right) \tag{5.43}$$

where: – $(\xi^\ell)_{\ell\ge1}$ is i.i.d., normally distributed (and comes from the Karhunen-Loève extension of W),

– the sequence $\left(t \mapsto \sqrt{\frac{2}{T}}\left(1-\cos\left(\frac{t}{\sqrt{\lambda_\ell}}\right)\right)\right)_{\ell\ge1}$ is *not* orthonormal in L^2_T.

In fact, the expansion (5.42) *converges* $\mathbb{P}$-$a.s.$ *and in* $L^1(\mathbb{P})$, *uniformly in* $t \in [0,T]$, since

$$\sup_{t\in[0,T]}\left|\sum_{\ell\ge1}\lambda_\ell\,\xi^\ell\sin^2\left(\frac{t}{2\sqrt{\lambda_\ell}}\right)\right| \le \sum_{\ell\ge1}\lambda_\ell|\xi^\ell|.$$

The series on the right hand of the inequality lies in $L^1(\mathbb{P})$ since $\sum_{\ell\geq 1}\lambda_\ell < +\infty$ and $\xi^\ell \sim \xi^1 \in L^1(\mathbb{P})$. The same holds for the integrated product quantizer expansion, that is

$$\widetilde{\int_0^{\cdot} W_s ds} := \int_0^{\cdot} \widehat{W}_s^{\sqrt{\lambda}\otimes x} ds = 2\sqrt{\frac{2}{T}} \sum_{\ell\geq 1} \lambda_\ell\, \widehat{\xi}^\ell \sin^2\left(\frac{t}{2\sqrt{\lambda_\ell}}\right) \tag{5.44}$$

since, by stationarity of the quantizer $x^{(\ell)}$ of ξ^ℓ, $\mathbb{E}|\widehat{\xi}^\ell| \leq \mathbb{E}|\xi^\ell|$ for every $\ell \geq 1$ (note that the $\mathbb{P}$-$a.s.$ convergence is trivial since $\widehat{\xi} = 0$ for ℓ large enough). One has to be aware that $\widetilde{\int_0^\cdot W_s ds}$ *is neither a product nor a Voronoi quantization* since it is defined on the Voronoi tessellation of the Brownian motion. For this very reason it is easy to compute and furthermore it satisfies a kind of stationary equation: one checks that $\sigma(\widetilde{\int_0^\cdot W_s ds}) = \sigma(\widehat{W}) = \sigma(\xi^\ell,\, \ell \geq 1)$ so that, $h \mapsto \int_0^\cdot h(s)ds$ being continuous and linear on L^2_T,

$$\mathbb{E}\left(\int_0^{\cdot} W_s ds \,|\, \widetilde{\int_0^{\cdot} W_s ds}\right) = \mathbb{E}\left(\int_0^{\cdot} W_s ds \,|\, \widehat{W}\right) = \widetilde{\int_0^{\cdot} W_s ds}.$$

Proposition 5 *Let $x^N \in \mathcal{O}_{pq}(N)$, $N \geq 1$. Set $\widetilde{\int_0^\cdot W_s ds}^{N} := \widetilde{\int_0^\cdot W_s ds}^{\sqrt{\lambda}\otimes x^N}$.*
(a) The quadratic quantization error is given by

$$\left\| \int_0^{\cdot} W_s ds - \widetilde{\int_0^{\cdot} W_s ds}^{N} \right\|_2^2 = 3\sum_{\ell\geq 1} \lambda_\ell^2 \left(1 - (-1)^{\ell-1}\frac{4\sqrt{\lambda_\ell}}{3T}\right) \min_{\mathbb{R}^{N_\ell}} D_{N_\ell}^{\mathcal{N}(0;1)}. \tag{5.45}$$

If $(\sqrt{\lambda}\otimes x^N)_{N\geq 1}$ is rate optimal for W then it is rate optimal for its antiderivative in the sense that

$$\left\| \int_0^{\cdot} W_s ds - \widetilde{\int_0^{\cdot} W_s ds}^{N} \right\|_2^2 = O((\log N)^{-\frac{3}{2}}).$$

(b) The $L^1(\mathbb{P})$-mean $\|\,.\,\|_{\sup}$-quantization error satisfies

$$\mathbb{E}\left(\sup_{t\in[0,T]} \left| \int_0^t W_s ds - \widetilde{\int_0^t W_s ds}^{N} \right| \right) \leq 2\sqrt{\frac{2}{T}} \sum_{\ell\geq 1} \lambda_\ell \min_{\mathbb{R}^{N_\ell}} \sqrt{D_{N_\ell}^{\mathcal{N}(0;1)}}. \tag{5.46}$$

Proof: (a) Temporarily set $E_\ell(t) = 1 - \cos\left(\frac{t}{\sqrt{\lambda_\ell}}\right)$. Then $|E_\ell|^2_{L^2_T} = T\left(\frac{3}{2} - 2(-1)^{\ell-1}\frac{\sqrt{\lambda_\ell}}{T}\right)$

and $$\left|\int_0^{\cdot} W_s ds - \widetilde{\int_0^{\cdot} W_s ds}^{N}\right|^2_{L^2_T} = \frac{2}{T}\sum_{\ell,m\geq 1}\lambda_\ell\lambda_m \mathbb{E}(\xi^\ell - \widehat{\xi}^\ell)(\xi^m - \widehat{\xi}^m)\,(E^\ell\,|\,E^m)_{L^2_T}$$

so that $$\left\|\left|\int_0^{\cdot} W_s ds - \widetilde{\int_0^{\cdot} W_s ds}^{N}\right|\right\|^2_2 = \frac{2}{T}\sum_{\ell\geq 1}\lambda_\ell^2\,\mathbb{E}(\xi^\ell - \widehat{\xi}^\ell)^2|E^\ell|^2_{L^2_T}.$$

This follows from the fact that the random variables $\xi^\ell - \widehat{\xi}^\ell, \ell \geq 1$, are independent and centered since $\mathbb{E}(\xi^\ell - \widehat{\xi}^\ell) = \mathbb{E}(\mathbb{E}(\xi^\ell|\widehat{\xi}^\ell) - \widehat{\xi}^\ell) = 0$. The rate of decay follows from the optimal size allocation in the right hand side of Inequality (5.46) which is standard (see [11]).

(b) easily follows from

$$\sup_{t\in[0,T]}\left|\int_0^t W_s ds - \widetilde{\int_0^t W_s ds}^{N}\right| = 2\sqrt{\frac{2}{T}}\sup_{t\in[0,T]}\left|\sum_{\ell\geq 1}\lambda_\ell(\xi^\ell - \widehat{\xi}^\ell)\sin^2\left(\frac{t}{2\sqrt{\lambda_\ell}}\right)\right| \leq 2\sqrt{\frac{2}{T}}\sum_{\ell\geq 1}\lambda_\ell\,|\xi^\ell - \widehat{\xi}^\ell|. \quad\diamond$$

Remarks. • One derives similarly from item (b) that the lowest $L^1(\mathbb{P})$-mean $L^\infty(dt)$-quantization error goes to zero at a $O\left((\log(N))^{-1}\right)$-rate.

• Some rates can be obtained for higher iterated integrals (and the Brownian bridge too).

6 The Romberg log-extrapolation

The aim of this paragraph is to propose a Romberg like extrapolation method to speed up the convergence of the quantization method. What follows is partially heuristic in that it relies on some claims on functional quantization which are still conjectures. For these reasons, we will focus on the Brownian motion and will not look for optimal assumptions.

Let $\Psi : (L^2_T, |\,.\,|_{L^2_T}) \to \mathbb{R}$ be a three times differentiable functional such that $\mathrm{D}^2\Psi$ and $\mathrm{D}^3\Psi$ are bounded. Let $(\chi^N)_{N\geq 1}$ denote a sequence of stationary rate optimal quantizers of the Brownian motion W (*e.g.* $\chi^N = \sqrt{\lambda}\otimes x^N_{\rm opt}$ in the K-L basis) and let $\widehat{W}^N$ denote their related quantizations.

It follows from the Taylor formula and Proposition 4 (stationarity property of $\widehat{W}^N$) that one can easily find $\zeta \in L^2_T$ such that

$$\mathbb{E}(\Psi(W)) = \mathbb{E}(\Psi(\widehat{W}^N)) + \frac{1}{2}\mathbb{E}(\mathrm{D}^2\Psi(\widehat{W}^N).(W-\widehat{W}^N)^{\otimes 2}) + \frac{1}{6}\mathbb{E}(\mathrm{D}^3\Psi(\zeta).(W-\widehat{W}^N)^{\otimes 3})$$

since $\mathbb{E}(\mathrm{D}\Psi(\widehat{W}^N).(W-\widehat{W}^N)) = \mathbb{E}(\mathrm{D}\Psi(\widehat{W}^N).(W-\mathbb{E}(W\,|\,\widehat{W}^N))) = 0$. Then, $\mathrm{D}^2\Psi$ being bounded and χ^N-rate optimal,

$$\mathbb{E}(\mathrm{D}^2\Psi(\widehat{W}^N).(W-\widehat{W}^N)^{\otimes 2}) = O\left((\log N)^{-1}\right),$$

but recent (finite dimensional) results (see [1], Theorem 6) suggest that, more precisely,

$$\mathbb{E}(\mathrm{D}^2\Psi(\widehat{W}^N).(W-\widehat{W}^N)^{\otimes 2}) = 2\,\kappa(\log N)^{-1} + o((\log N)^{-1}) \quad \text{as} \quad N\to\infty$$

where $\kappa > 0$ is real constant. Moreover, still relying on [1] (Proposition 1), one shows that

$$\mathbb{E}|W-\widehat{W}^N|^3_{L^2_T} = O((\log N)^{-\frac{3}{2}+\eta}), \qquad \forall\,\eta>0.$$

Then, a Romberg like speeding up procedure can be implemented as follows: one computes $\mathbb{E}(\Psi(\widehat{W}^M))$ and $\mathbb{E}(\Psi(\widehat{W}^N))$, $M<N$, $M \asymp N^r$, $r\in(0,1)$. Solving the linear system

$$\mathbb{E}(\Psi(W)) = \mathbb{E}(\Psi(\widehat{W}^M)) + \frac{\kappa}{\log M} + O((\log M)^{-\frac{3}{2}+\eta}),\ \mathbb{E}(\Psi(W)) = \mathbb{E}(\Psi(\widehat{W}^N)) + \frac{\kappa}{\log N} + O((\log N)^{-\frac{3}{2}+\eta})$$

yields the announced log-extrapolation formula

$$\mathbb{E}(\Psi(W)) = \frac{\log N\times\mathbb{E}(\Psi(\widehat{W}^N)) - \log M\times\mathbb{E}(\Psi(\widehat{W}^M))}{\log N - \log M} + O\left((\log N)^{-\frac{3}{2}+\eta}\right),\ \forall\,\eta>0. \tag{6.47}$$

So we passed from a $O((\log N)^{-1})$-rate to an (at least) $O\left((\log N)^{-\frac{3}{2}+\eta}\right)$-rate.

7 Pricing derivatives using functional quantization

7.1 Pricing Asian options in a Black-Scholes model

One considers a Black-Scholes dynamics with maturity T,

$$dS_t = S_t\,(r dt + \sigma\, dW_t), \quad S_0 = s_0 > 0 \qquad (r>0).$$

The premium of an Asian (European) Call option with strike price K is given by

$$C_{as}(s_0,K) \;:=\; e^{-rT}\mathbb{E}\left(\left(\frac{1}{T}\int_0^T S_t dt - K\right)_+\right) = e^{-rT}\mathbb{E}\left(\left(s_0\frac{1}{T}\int_0^T e^{\sigma W_t+(r-\sigma^2/2)t}dt - K\right)_+\right)$$

where $x := \max(x,0)$ denotes the nonnegative part of the real number x. We want to approximate $C_{as}(s_0,K)$ using quadratic functional quantization.

(a) "Crude" functional quantization method: One simply computes

$$\widehat{C}_{as}(s_0,K,\chi^N) := e^{-rT}\mathbb{E}\left(\left(s_0\frac{1}{T}\int_0^T e^{\sigma \widehat{W}_t^N+(r-\sigma^2/2)t}dt - K\right)_+\right)$$

where the process $(\widehat{W}_t^N)_{t\in[0,T]}$ denotes the quantization of the Brownian motion W by the product quantizer $\chi^N := \sqrt{\lambda}\otimes x^N_{\rm rec}$ which induces the lowest quantization error among all (Karhunen-Loève) product quantizers having at most N-components. The functional

$$\omega \mapsto \left(\frac{s_0}{T}\int_0^T e^{\sigma\omega(t)+(r-\sigma^2/2)t}dt - K\right)_+$$

is convex, consequently combining (4.16) and the stationarity of $(\widehat{W}_t)_{t\in[0,T]}$

$$\widehat{C}_{as}(s_0,K,\chi^N) \le C_{as}(s_0,K). \tag{7.48}$$

This bound can be improved by considering *any $\sqrt{\lambda}$-scaled product quantizer $\sqrt{\lambda}\otimes x$ of W, with $x\in\mathcal{O}_{pq}(N)$*. Then, with obvious notations

$$\sup_{x\in\mathcal{O}_{pq}(N)} \widehat{C}_{as}(s_0,K,\sqrt{\lambda}\otimes x) \le C_{as}(s_0,K).$$

Furthermore, it follows from (4.13) and (4.18) that, for any $x\in\mathcal{O}_{pq}(N)$,

$$\begin{aligned} 0\le C_{as}(s_0,K)-\widehat{C}_{as}(s_0,K,\sqrt{\lambda}\otimes x) &\le 2\, s_0\sigma e^{-rT+(r-\frac{\sigma^2}{2})_+T}\,\mathbb{E}(|\exp(\sigma W)|_{L^2_T})\,\|W-\widehat{W}^{\sqrt{\lambda}\otimes x}\|_2 \\ &\le 2\, s_0 e^{-(r\wedge\frac{\sigma^2}{2})T}(e^{\sigma^2 T}-1)^{\frac12}\|W-\widehat{W}^{\sqrt{\lambda}\otimes x}\|_2. \end{aligned}$$

In particular

$$0\le C_{as}(s_0,K)-\widehat{C}_{as}(s_0,K,\chi^N)\le 2\, s_0 e^{-(r\wedge\frac{\sigma^2}{2})T}(e^{\sigma^2T}-1)^{\frac12}\min_{x\in\mathcal{O}_{pq}(N)}\|W-\widehat{W}^{\sqrt{\lambda}\otimes x}\|_2 = O((\log N)^{-\frac12})$$

as $N\to\infty$. One also has, using (4.17), that

$$|\mathbb{E}\,\widehat{S}_{T}^{\chi^N}-\mathbb{E}\,S_{T}|=o((\log N)^{-\frac{1-\varepsilon}{2}})\quad\text{as}\quad N\to+\infty\quad\text{for every }\varepsilon>0.$$

(b) Functional quantization and geometric control variate variable: The standard Kemna-Vorst control variate variable for Asian options is "suggested" by Jensen Inequality

$$\left(s_0\frac{1}{T}\int_0^T e^{\sigma W_t+(r-\sigma^2/2)t}dt-K\right)_+\geq\left(s_0 e^{\frac{1}{T}\int_0^T(\sigma W_t+(r-\sigma^2/2)t)dt}-K\right)_+=\left(s_0 e^{(r-\frac{\sigma^2}{2})\frac{T}{2}+\frac{\sigma}{T}\int_0^T W_t dt}-K\right)_+.$$

The contingent claim on the right hand side of the inequality is called a *geometric Asian Call* since it is a geometric mean (see also [8]). Using that $\frac{1}{T}\int_0^T W_t dt\sim\mathcal{N}(0;T^2/3)$, one shows that the Black-Scholes premium $C_{geo}(s_0,K)$ for the geometric Asian Call is given by

$$C_{geo}(s_0,K)=C_{BS}\left(s_0\,e^{-(\frac{\sigma^2}{12}+\frac{r}{2})T},K,r,\sigma/\sqrt{3},t\right)$$

where $C_{BS}(x,y,r,\sigma,\theta)$ stands for the usual Black & Scholes formula for the Call option where x denotes the spot, y the strike, r the interest rate, σ the constant volatility and θ the distance to maturity. Let us note that the former price can be computed by quantization by

$$\widehat{C}_{geo}(s_0,K,\chi^N)=e^{-rT}\mathbb{E}\left(\left(s_0\,e^{(r-\frac{\sigma^2}{2})\frac{T}{2}+\frac{\sigma}{T}\widetilde{\int_0^T W_t dt}^{N}}-K\right)_+\right)$$

where $\widetilde{\int_0^T W_t dt}^{N}$ is given by (5.44). So, one can set (with cv for Control Variate)

$$\widehat{C}_{as}^{cv}(s_0,K,\chi^N)=e^{-rT}\mathbb{E}\left(\left(\frac{s_0}{T}\int_0^T e^{\sigma\widehat{W}_t^N+(r-\sigma^2/2)t}dt-K\right)_+-\left(s_0 e^{(r-\frac{\sigma^2}{2})\frac{T}{2}+\frac{\sigma}{T}\widetilde{\int_0^T W_t dt}^{N}}-K\right)_+\right)+C_{geo}(s_0,K).$$

Let us note that we use here the term "control variate" in reference to the variance reduction in a Monte Carlo method. The geometric Asian is not used here in a Monte Carlo framework and the expectation of the previous equation is computed by quantization. However, we will see on the numerical results that it seems to play the same rôle as a control variate. Note also that the computation of $\widetilde{\int_0^T W_t dt}^{N}$ is based on (5.44) and needs no time discretization of the integral. One defines similarly a geometric Asian Put which can be used as a control variate for the Asian Put (note that it is an upper-bound of the regular Asian put).

(c) FUNCTIONAL QUANTIZATION AND PARITY: It is also possible to use the following Call-Put parity

$$C_{as}(s_0,K) - P_{as}(s_0,K) = e^{-rT}s_0\frac{1}{T}\int_0^T \mathbb{E}(e^{\sigma W_t+(r-\sigma^2/2)t})dt - Ke^{-rT} = s_0\frac{1-e^{-rT}}{rT} - Ke^{-rT}$$

to compute the Call premium by parity (par) *i.e.*

$$\widehat{C}^{par}_{as}(s_0,K,\chi^N) = s_0\frac{1-e^{-rT}}{rT} - Ke^{-rT} + \widehat{P}_{as}(s_0,K,\chi^N). \tag{7.49}$$

(Note that the Put contingent claim is not a convex functional of the generic Brownian path so that no inequality such as (7.48) holds for $\widehat{C}^{par}_{as}(s_0,K,\chi^N)$). The main interest is that $\widehat{P}_{as}(s_0,K,\chi^N)$ is often lower than $\widehat{C}_{as}(s_0,K,\chi^N)$ so that the same relative error on the Put has less impact.

Numerical experiments and results: As a first step, four time discretization methods have been tested to compute the integral $\int_0^T s_0 \exp\left(\sigma\widehat{W}_t - \frac{\sigma^2 t}{2}\right)dt$. Two are second order methods: the trapezoid method, the *midpoint method* (corresponding to the optimal quantization of the uniform distribution over the interval $[0,T]$) *i.e.*

$$\int_0^T g(t)dt \approx \frac{1}{n}\sum_{k=1}^n g(t_k), \qquad t_k = \frac{(2k-1)}{2n}T,\ k=1,\ldots,n.$$

and two are fourth order methods: the fourth-order Runge-Kuta method and the fourth order midpoint method, based on the introduction of the (weighted) derivatives of g at points t_k.

We specified numerical values for the parameters following the tests carried out in [10], *i.e.*

$$s_0 = K = 100, \qquad T=1, \qquad r=10\%, \qquad \sigma = 20\%.$$

The reference values for the Asian Call and Put are, still following [10]

Asian Call Premium = 7.041   Theoretical Asian Put Premium = 2.362

The best compromise between complexity and efficiency is the second order midpoint method (see Figure 4). To reach this conclusion we let n become large for two values of N (using the crude functional quantization method (FQ) for the Asian Call). Let us note that the “limiting value" (when the time discretization step $1/n$ goes to 0) looks far the true value (≈ 7.041) with all the four methods. This comes from the scale specification since our aim in this first experiment is exclusively to compare the different *time* integration methods. In fact, it will be seen below that, for a given couple (n, N) of time-space discretization parameters, best prices are usually obtained using the Call-Put parity.

Now, we come to testing the FQ method itself. Asian option premia are computed using $N = 96$ (Table 3), $N = 966$ (Table 4) and $N = 9\,984$ (Table 5). Time integration is performed by the midpoint scheme with $n = 20$. For each values of N, the second row displays $2 \times Std_{_N}$ where $Std_{_N}$ denotes the *relative* standard deviation of the corresponding Monte Carlo N-estimator which defines its 95.5%-confidence interval. This estimator is based on the simulation of the K-L expansion (5.36) of the Brownian motion (truncated at $\ell = 100$). The third row shows the decomposition producing the lowest distortion and its value. Then, the successive rows give the results for different factorizations of N. The third (resp. fourth) column gives the Call premium $\widehat{C}_{as}$ by "crude" FQ (resp. $\widehat{C}^{cv}_{as}$ with the Kemna-Vorst control variate). The fifth (resp. sixth) one displays the Put premium $\widehat{P}_{as}$ obtained by "crude" FQ (resp. $\widehat{P}^{cv}_{as}$ with the control variate). The seventh (resp. eighth) one displays the Call premium obtained using the Call-Put parity relation (7.49) and the fifth column (resp. using (7.49) and the sixth column).

The results in Tables 3, 4, 5 are in a descending order *with respect to the Call premia obtained by crude FQ* (column 3). This is justified by the fact that this method always produces a lower bound for the premium (although this sorting is unrealistic in practice).

For both the decomposition with the lowest distortion ($\star$) – the one of interest for applications – and the one with the highest premium $\widehat{C}_{as}$ for the Call (fourth row of every table), the relative error is added between brackets.

At this stage, the tables suggest that the most performing method to compute the Call (at-the-money) is the computation by parity, from the Put computed by "crude" FQ using the "record" scaled product quantizer χ^N (decomposition $\star$, see column 7 in Tables 3, 4, 5): the relative error is always less than 0.2% (but one loses the lower bound property). as $N \geq 1\,000$. Within this range of values of N ($N \leq 10\,000$) the computation is instantaneous. Let us note that the Kemna-Vorst control variate variable (columns 4, 6, 8) seems globally less efficient than in the Monte Carlo method but give errors which are of the same order than the relative standart deviations of the Monte-Carlo estimators based on K-L expansion (second line) at least for $N = 96$ and $N = 966$. This tell us that this control variate variable seems to play the same rôle both in the functional quantization and Monte Carlo method when N is small. When N becomes larger, this control variate method seems more efficient in the Monte Carlo method (based on the K-L expansion (5.36)).

$N = 96$	n	$\widehat{C}_{as}$	$\widehat{C}^{cv}_{as}$	$\widehat{P}_{as}$	$\widehat{P}^{cv}_{as}$	$\widehat{C}^{par}_{as}$	$\widehat{C}^{par\&cv}_{as}$	Decomposition
$2 \times Std_{96}$	20	(24.49%)	(1.14%)	(37.11%)	(1.33%)	(12.45%)	(0.47%)	-
96	20	Best decomp($\star$): $96 = 12 \times 4 \times 2$, Distor* $=$ **0.051276**						
96	20	**6.957** (**1.2%**)	**6.971** (**1.0%**)	**2.387** (**1.1%**)	**2.396** (**1.4%**)	**7.066** (**0.3%**)	**7.075** (**0.5%**)	$24 - 4$
96	20	6.957	6.981	2.377	2.393	7.056	7.071	$16 - 3 - 2$
96	20	**6.953** (**1.3%**)	**6.983** (**0.8%**)	**2.372** (**0.4%**)	**2.390** (**1.2%**)	**7.051** (**0.1%**)	**7.069** (**0.4%**)	$\mathbf{12 - 4 - 2} \star$
96	20	6.952	6.969	2.386	2.399	7.0651	7.078	$32 - 3$
96	20	6.951	6.975	2.379	2.396	7.0582	7.074	$16 - 6$
96	20	6.950	6.972	2.381	2.396	7.0595	7.075	$24 - 2 - 2$
96	...	...	...	etc	...	...	...	...

Table 3. *Asian Call approximations for $N = 96$*

$N = 966$	n	$\widehat{C}_{as}$	$\widehat{C}^{cv}_{as}$	$\widehat{P}_{as}$	$\widehat{P}^{cv}_{as}$	$\widehat{C}^{par}_{as}$	$\widehat{C}^{par\&cv}_{as}$	Decomposition
$2 \times Std_{966}$	20	(7.72%)	(0.36%)	(11.70%)	(0.42%)	(3.93%)	(0.14%)	-
966	20	Best decomp($\star$): $966 = 23 \times 7 \times 3 \times 2$, Distor* $=$ **0.035195**						
966	20	**6.988** (**0.7%**)	**6.999** (**0.5%**)	**2.377** (**0.3%**)	**2.383** (**0.3%**)	**7.055** (**0.2%**)	**7.062** (**0.3%**)	$\mathbf{23 - 7 - 3 - 2} \star$
966	20	6.987	6.992	2.382	2.386	7.061	7.065	$46 - 7 - 3$
966	20	6.984	6.995	2.378	2.385	7.057	7.064	$23 - 7 - 6$
966	20	6.984	6.989	2.384	2.388	7.063	7.067	$69 - 7 - 2$
966	20	6.983	6.994	2.379	2.386	7.058	7.065	$23 - 14 - 3$
966	20	6.980	6.990	2.380	2.387	7.059	7.066	$23 - 21 - 2$
966	20	6.971	6.977	2.388	2.393	7.067	7.072	$69 - 14$
966	20	6.970	6.978	2.388	2.393	7.066	7.072	$46 - 21$
966	20	6.970	6.978	2.387	2.393	7.066	7.072	$42 - 23$

Table 4. *Asian Call approximations for $N = 966$*

N	n	$\widehat{C}_{as}$	$\widehat{C}^{cv}_{as}$	$\widehat{P}_{as}$	$\widehat{P}^{cv}_{as}$	$\widehat{C}^{par}_{as}$	$\widehat{C}^{par\&cv}_{as}$	Decomposition
$2 \times Std_{9984}$	20	(2.40%)	(0.11%)	(3.64%)	(0.13%)	(1.22%)	(0.04%)	-
9 984	20	Best decomp.($\star$): 9 984 = $26{\times}8{\times}4{\times}3{\times}2{\times}2$, Distor* = **0.026435**						
9 984	20	**7.004** **(0.5%)**	**7.007** **(0.4%)**	**2.377** **(0.6%)**	**2.380** **(0.7%)**	**7.056** **(0.2%)**	**7.058** **(0.2%)**	$52-8-4-3-2$
9 984	20	7.003	7.008	2.376	2.379	7.055	7.058	$39-8-4-2-2-2$
9 984	20	7.003	7.008	2.376	2.379	7.055	7.058	$39-8-4-4-2$
9 984	20	7.003	7.006	2.378	2.380	7.057	7.059	$52-12-4-2-2$
9 984	20	7.003	7.007	2.377	2.380	7.056	7.059	$52-8-6-2-2$
9 984	20	7.003	7.006	2.378	2.380	7.056	7.059	$48-13-4-2-2$
9 984	...	...	...	etc	...	...	...	...
9 984	20	7.002	7.005	2.378	2.380	7.057	7.059	$52-16-3-2-2$
9 984	20	**7.002** **(0.5%)**	**7.010** **(0.5%)**	**2.373** **(0.4%)**	**2.378** **(0.3%)**	**7.052** **(0.1%)**	**7.057** **(0.3%)**	$\mathbf{26-8-4-3-2-2}$ $\star$
9 984	...	...	...	etc	...	...	...	...

Table 5. *Asian Call approximations for* $N = 9\,984$

Further values reported in Table 6 below were obtained using the N_ℓ given by (5.27) (in fact their upper truncation) so they are *a priori* not even optimal among all the decompositions of their product $N = N_1 \cdots N_{m_N}$. This confirms that the convergence as N grows is slow. The quite good results

N	n	$\widehat{C}_{as}$	$\widehat{C}_{as}^{cv}$	$\widehat{P}_{as}$	$\widehat{P}_{as}^{cv}$	$\widehat{C}_{as}^{par}$	$\widehat{C}_{as}^{par\&cv}$	Decomposition
13 500	20	7.002	7.010	2.373	2.378	7.052	7.057	$25-9-5-4-3$
		(0.5%)	(0.4%)	(0.5%)	(0.7%)	(0.06%)	(0.2%)	
40 500	20	7.005	7.013	2.372	2.377	7.050	7.055	$25-9-5-4-3-3$
		(0.5%)	(0.4%)	(0.5%)	(0.7%)	(0.03%)	(0.2%)	
104 400	20	7.009	7.015	2.372	2.376	7.051	7.055	$29-10-6-5-4-3$
		(0.5%)	(0.4%)	(0.5%)	(0.7%)	(0.04%)	(0.2%)	
313 200	20	7.012	7.018	2.371	2.375	7.050	7.054	$29-10-6-5-4-3-3$
		(0.4%)	(0.3%)	(0.4%)	(0.5%)	(0.03%)	(0.2%)	

Table 6. *Asian Call approximations for larger values of N*

obtained for small values of N is an important asset of functional quantization.

THE ROMBERG log-EXTRAPOLATION. Although the regularity assumptions are clearly not fulfilled by the Asian payoff, we tested the Romberg log-extrapolation (6.47) and reported in Table 7 below the results obtained with $N = 966$ and $M = 9\,984$ by using the record product quantizers. It turns out that this drastically improves all the approaches in such a way which seems not to depend on the approach itself ("crude" Functional Quantization, control variate variable, by parity, and so on). This suggests that a functional quantization approach including a Romberg log-extrapolation provides an extremently efficient numerical method for this problem.

Romberg[966-9984]	n	$\widehat{C}_{as}$	$\widehat{C}_{as}^{cv}$	$\widehat{P}_{as}$	$\widehat{P}_{as}^{cv}$	$\widehat{C}_{as}^{par}$	$\widehat{C}_{as}^{par\&cv}$
	20	7.041	7.041	2.364	2.364	7.042	7.042
		(0.00%)	(0.00%)	(0.08%)	(0.08%)	(0.01%)	(0.01%)

Table 7. *Romberg[966-9984]* log-*extrapolation using record product quantizers*

7.2 Pricing vanilla options in a Heston model

In this paragraph, we consider a Heston stochastic volatility model for the dynamics of an asset price process:

$$\begin{aligned} dS_t &= S_t(r\,dt + \sqrt{v_t})dW_t^1, \qquad S_0 = s_0 > 0, && (7.50)\\ *[.4em]dv_t &= k(a - v_t)dt + \vartheta\sqrt{v_t}dW_t^2 \quad v_0 > 0, \qquad \text{with} \qquad <W^1, W^2>_t = \rho\, t,\ \rho \in [-1, 1], \end{aligned}$$

where r denotes the (constant) interest rate and (v_t) denotes the square stochastic volatility process and a, k, ϑ are non-negative real parameters. This model was introduced by Heston in 1993 (see [6]). The equation for the (v_t) has a unique (strong) pathwise continuous solution living in $\mathbb{R}_+$ (see *e.g.* [9] and [7], p.235). Thereis a semi-closed form for vanilla European Call and Put options based on some integrals of the characteristic function for which a closed form is available (see [9]). We will use it as a reference for our experiments. Our aim is to price by functional quantization (at time 0) European Calls (and Puts) on the underlying asset (S_t) with strike price K and maturity $T > 0$, *i.e.*

$$\mathrm{Call}_{Hest}(S_0, K, r) = e^{-rT}\mathbb{E}((S_T - K)_+) \qquad \text{and} \qquad \mathrm{Put}_{Hest}(S_0, K, r) = e^{-rT}\mathbb{E}((K - S_T)_+).$$

As a first step, we follow an approach which works for more general dynamics of the stochastic volatility. First we project W^1 onto W^2 so that

$$W_t^1 = \rho W_t^2 + \sqrt{1-\rho^2}\,\widetilde{W}_t^1,$$

with $\widetilde{W}^1$ a standard Brownian motion independent of W^2. Itô calculus shows that

$$S_t = s_0 \exp\left(-\frac{\rho^2}{2}\bar{v}_t\, t + \rho \int_0^t \sqrt{v_s} dW_s^2\right) \exp\left((r - \frac{1-\rho^2}{2}\bar{v}_t)t + \sqrt{1-\rho^2}\int_0^t \sqrt{v_s} d\widetilde{W}_s^1\right)$$

with $\bar{v}_t = \frac{1}{t}\int_0^t v_s ds$. Consequently, using the independence of $\widetilde{W}^1$ and W^2, one derives that

$$\mathrm{Call}_{Hest}(S_0, K, T, v_0, r) \;=\; \mathbb{E}\left(e^{-rT}\mathbb{E}\left((S_T - K)_+ \,|\, \mathcal{F}_T^{W^2}\right)\right) = \mathbb{E}\left(\mathrm{Call}_{BS}\left(S_0^{(v)}, K, T, \left((1-\rho^2)\bar{v}_T\right)^{\frac{1}{2}}, r\right.\right.$$

$*[.4em]$with $$S_0^{(v)} \;=\; s_0 \exp\left(-\frac{\rho^2}{2}\bar{v}_T T + \rho \int_0^T \sqrt{v_s} dW_s^2\right)$$

where $\mathrm{Call}_{BS}(s_0, K, T, \sigma, r)$ denotes the regular (r, σ, T)-Black-Scholes model premium function. Then the specific dynamics of (v_t) yields([1])

$$\int_0^t \sqrt{v_s} dW_s^2 = \frac{v_t - v_0 - kat + k\int_0^t v_s ds}{\vartheta}$$

so that finally

$$\mathrm{Call}_{Hest}(S_0, K, T, v_0, r) \;=\; \mathbb{E}\,\Phi_c(\rho(v_T - v_0), \bar{v}_T) \tag{7.51}$$

[1] The key point in what follows is to express the stochastic integral $\int_0^t \sqrt{v_s} dW_s^2$ as a functional of v_t, v_0 and an integral functional of (v_s). If the variance process follows a general diffusion process $dv_t = b(v_t)dt + \vartheta(v_t)dW_t^2$ then one may apply under appropriate regularity assumption, Itô's formula to the function $\varphi(v) := \sqrt{v}/\vartheta(v)$ to get such an expression.

$$\text{with}\qquad \Phi_c(v,\bar v) = \mathrm{Call}_{BS}\left(s_0 \exp\left(-\rho\left(\frac{ka}{\vartheta} - (\frac{k}{\vartheta} - \frac{\rho}{2})\bar v\right) T + \frac{v}{\vartheta}\right), K, T, \left((1-\rho^2)\bar v\right)^{\frac{1}{2}}, r\right).$$

An analogous formula holds for $\mathrm{Put}_{Hest}(S_0, K, T, v_0, r)$ by replacing *mutatis mutandis* Call_{BS} by Put_{BS} in (7.51). Note that when $\rho = 0$, (7.51) only depends on the L^2-continuous linear functional $\bar v_T$.

7.2.1 The quantization procedure

A first way to numerically quantize (v_t) is to follow – at least formally – the approach developed in [13] to quantize Brownian diffusions with Lipschitz continuous coefficients. One considers again the sequence $\chi^N := \sqrt{\lambda} \otimes x^N_{\mathrm{rec}}$, $N \geq 1$, of record product quantizers of the standard Brownian motion (which are explicit C^∞ functions). For convenience, we will consider now the "record" subsequence *i.e.* assume that $N = N_{\mathrm{rec}}$.

Assume for a while that (v_t) is a generic Brownian diffusion

$$dv_t = b(v_t)dt + \vartheta(v_t)dW^2_t \qquad (\vartheta \geq 0)$$

Some quantizers for (v_t) can be designed from the sequence (χ^N) as follows: one introduces the Lamperti transform of the diffusion defined by $L(v) := \int_0^v \frac{dv}{\vartheta(v)}$ (assumed to be real-valued and increasing). Then, $U_t := L(v_t)$ satisfies is solution of the SDE

$$dU_t = \beta(U_t)dt + dW_t$$

with a linear Brownian perturbation term. Then, one defines, a N-quantizer of (v_t) by setting

$$y^N_i(t) = L^{-1}(u^N_i(t)) \quad \text{where} \quad u^N_i(t) = L(v_0) + \int_0^t \beta(u^N_i(t))dt + \chi^N_i(t), \quad i = 1, \ldots, N.$$

Elementary computations show that $y^N = (y^N_i)_{1 \leq i \leq N}$ is solution of the system of integral equations

$$y^N_i(t) = v_0 + \int_0^t [b(y^N_i(s)) - \frac{1}{2}\vartheta\vartheta'(y^N_i(s))]ds + \int_0^t \vartheta(y^N_i(s))d\chi^N_i(s),\ i = 1, \ldots, N. \quad (7.52)$$

When b and ϑ are Lipschitz continuous, it is established in [13] (Theorem 1 and the following Application) that if the sequence $(\chi^N)_{N\geq 1}$ is rate optimal in $L^2_{L^2_T}(\Omega, \mathbb{P})$ for the Wiener measure, then $(y^N)_{N\geq 1}$ is rate optimal in $L^p_{L^2_T}(\Omega,\mathbb{P})$ for every $p\in[1,2)$. More precisely, it is established in [13] that the sequence of non-Voronoi N-quantizations

$$\widetilde{v}^N_t = \sum_{1\leq i\leq N} y^N_i(t)\mathbf{1}_{C_i(\chi^N)}(W^2),\ N\geq 1,$$

satisfies $\|\,|v-\widetilde{v}^N|_{L^2_T}\|_p = O((\log N)^{-1/2})$, $p\in[1,2)$. When $p=2$ a straightforward adaptation of the proof yields a $O((\log N)^{-\frac{1}{2}+\varepsilon})$-rate in $L^p_{L^2_T}(\Omega,\mathbb{P})$ for every $\varepsilon>0$. The quantization $\widetilde{v}^N_t$ is not Voronoi since it is defined on the Voronoi tessellation of W^2, but its distribution is given by the $\mathbb{P}_W$-weights of the cells $C_i(\chi^N)$ which are known by (5.38). In our non-Lipschitz setting ($b(v)=-k(v-a)$, $\vartheta(v)=\vartheta\sqrt{v}$), y^N satisfies

$$y^N_i(t) = v_0 + k\int_0^t\left(a-\frac{\vartheta^2}{4k}-y^N_i(s)\right)ds + \vartheta\int_0^t\sqrt{y^N_i(s)}d\chi^N_i(s), \qquad i=1,\ldots,N. \quad (7.53)$$

If $a>\vartheta^2/(4k)$, any solution of (7.53) is positive and, once again a simple adaptation of the proof of Theorem 1 in [13] shows that $\|\,|v-\widetilde{v}^N|_{L^2_T}\|_2 = O((\log N)^{-\frac{1}{2}+\varepsilon})$. Numerical implementation of this functional quantization method simply needs to use a discretization scheme of (7.53) like the Euler, the midpoint or the Runge-Kuta schemes.

However, in view of investigating the efficiency of functional quantization, this approach suffers from mixing two kinds of error: one due to the discretization scheme of the integral equation system and one due to functional quantization. So, to be more illustrative of the numerical performances of functional quantization, we will consider the Heston model in the case $a=\frac{\vartheta^2}{4k}$ since, as noticed by Rogers in [19], one may assume without loss of generality that the process (v_t) is the square of a scalar Ornstein-Uhlenbeck process

$$dX_t = -\frac{k}{2}X_t dt + \frac{\vartheta}{2}dW^2_t, \qquad X_0=\sqrt{v_0}. \quad (7.54)$$

Having in mind that the N-quantizers χ^N given by (5.37) read

$$\chi^N_{\underline{i}}(t) = \sqrt{\frac{2}{T}}\sum_{\ell\geq 1} x^{(N_\ell)}_{i_\ell}\frac{T}{\pi(\ell-1/2)}\sin\left(\pi(\ell-1/2)\frac{t}{T}\right), \qquad \underline{i}=(i_1,\ldots,i_\ell,\ldots)\in\prod_{\ell\geq 1}\{1,\ldots,N_\ell\},$$

where $x=\prod_{\ell\geq 1}x^{(\ell)}\in\mathcal{O}_{pq}$ the solutions of the integral system (7.52) associated to X

$$x_{\underline{i}}(t) = \sqrt{v_0} - \frac{k}{2}\int_0^t x_{\underline{i}}(s)\,ds + \frac{\vartheta}{2}\chi^N_{\underline{i}}(t), \quad i=1,\ldots,N \quad (7.55)$$

are given for every $\underline{i} = (i_1, \ldots, i_\ell, \ldots) \in \prod_{\ell\geq 1}\{1, \ldots, N_\ell\}$ by

$$x^N_{\underline{i}}(t) = e^{-kt/2}\sqrt{v_0} + \frac{\vartheta}{2}\sum_{\ell\geq 1} x^{(\ell)}_{i_\ell}\,\widetilde{c}_\ell\,\varphi_\ell(t) \qquad \text{with } \widetilde{c}_\ell := \frac{T^2}{(\pi(\ell-1/2))^2+(kT/2)^2}$$

and $$\varphi_\ell(t) := \sqrt{\frac{2}{T}}\left(\frac{\pi}{T}(\ell-1/2)\sin\left(\pi(\ell-1/2)\frac{t}{T}\right)+\frac{k}{2}\left(\cos\left(\pi(\ell-1/2)\frac{t}{T}\right)-e^{-kt/2}\right)\right).$$

This time, still following [13], we have for every $p\in[1,2)$,

$$\|\widetilde{X}^N - X\|_p \leq C_{p,k,\vartheta,T}\|\widehat{W^2}^{\chi^N} - W^2\|_2 = O\left((\log N)^{-\frac{1}{2}}\right) \tag{7.56}$$

where $\widetilde{X}^N$ is the non-Voronoi quantization defined by

$$\widetilde{X}^N_t = \sum_{i=1}^{N} x^N_{\underline{i}}(t)\mathbf{1}_{C_{\underline{i}}(\chi^N)}(W^2) = e^{-kt/2}\sqrt{v_0} + \frac{\vartheta}{2}\sum_{\ell\geq 1}\widehat{\xi}^{x(\ell)}_\ell\widetilde{c}_\ell\,\varphi_\ell(t), \quad t\in[0,T].$$

One designs a (non-Voronoi) N-quantization for the process (v_t) by setting

$$\widetilde{v}^N_t = (\widetilde{X}^N_t)^2 = \sum_{\underline{i}}(x^N_{\underline{i}}(t))^2\mathbf{1}_{C_{\underline{i}}(\chi^N)}(W^2). \tag{7.57}$$

Then, one derives from (7.56) that, for every $p\in[1,2]$,

$$\|\,|\widetilde{v}^N - v|_{L^2_T}\|_p = O\left((\log N)^{-(\frac{1}{2}-\varepsilon)}\right) \quad \text{for every } \varepsilon > 0. \tag{7.58}$$

Finally, in practice, one computes $\mathrm{Call}_{Hest}(S_0, K, T, r)$ by

$$\mathrm{Call}_{Hest}(s_0, K, T, v_0, r) = \mathbb{E}(\Phi_c(\rho(v_T - v_0), \bar{v}_T)) \tag{7.59}$$
$$*[.4em] \approx \mathbb{E}(\Phi_c(\rho(\widetilde{v}_T - v_0), \overline{\widetilde{v}}_T))$$
$$*[.4em] = \sum_{\underline{i}}\Phi_c\left(\rho\left((x^N_{\underline{i}})^2(T) - v_0\right), \overline{(x^N_{\underline{i}})^2}(T)\right)\mathbb{P}(\widehat{W}^2 = \chi^N_{\underline{i}}) \tag{7.60}$$

where the probability distribution $(\mathbb{P}(\widehat{W}^2 = \chi^N_{\underline{i}}))_{\underline{i}}$ is given by (5.38). When $\rho \neq 0$, no simple error bound is available since we do not know the rate of pointwise quantization of v_T by quadratic functional quantizers.

When $\rho = 0$, $\bar{v} \mapsto \Phi_c(0, v_0, \bar{v})$ is clearly Lipschitz, so a $O\left((\log N)^{-(\frac{1}{2}-\varepsilon)}\right)$-rate holds in (7.60). Furthermore, functions $\sigma \mapsto \mathrm{Put}_{BS}(s_0, K, T, \sigma, r)$ and its Call counterpart are infinitely differentiable on $(0, +\infty)$ and $u \mapsto \bar{u}_{_T} := \frac{1}{T}\int_0^T u(s)ds$ is an L^2_T-continuous linear functional. On the other hand, the solution of the integral equation $x(t) = x(0) - \frac{k}{2}\int_0^t x(s)ds + \frac{\vartheta}{2}\,\xi(t)$ is also an L^2_T-continuous linear functional functional of ξ. Consequently, one may write (7.59)

$$\mathrm{Put}_{Hest}(s_0, K, T, r) = \mathbb{E}(\Psi_p(W^2)) \quad \text{and} \quad \mathrm{Call}_{Hest}(s_0, K, T, r) = \mathbb{E}(\Psi_c(W^2))$$

where Ψ_p and Ψ_c are infinitely differentiable and Ψ_p is bounded with all its differentials .

7.2.2 Numerical experiments and results

All the formulæ for the Call can be straightforwardly adapted to the Put. Furthermore, the Call-Put parity holds which is a second way to compute the Call premium which has a lower variance. We used as a reference price the closed form available for the Heston model (approximate accuracy 10^{-2}).

Following the results of the experiments carried out with the Asian option we compute time integrals by the midpoint method with $n = 20$, *i.e.*

$$\overline{(x^N_{\underline{i}})^2} = \frac{1}{T}\int_0^T (x^N_{\underline{i}}(s))^2 ds \approx \frac{1}{n}\sum_{k=1}^{n}(x^N_{\underline{i}}(t_k))^2 \quad \text{with} \quad t_k = \frac{(2k-1)T}{2n}.$$

with the lowest quadratic quantization error as theywere computed The parameters of the Heston model are specified as follows

$$s_0 = 50, \quad r = 0.05, \quad T = 1, \quad \rho = 0.5, \quad v_0 = a = 0.01, \quad \vartheta = 0.1, \quad k = 0.25$$

so that $\vartheta^2 = 4\,k\,a$ (and $\mathbb{E}\,v_t = a$, $t \in [0, T]$). In this section, we carried out our numerical experiments on a whole vector of strike prices, K running from 44 up to 56 (with step 1) to evaluate the performances of the method for in-the-money, at-the-money and out-of-the-money options. The premia of these Heston Call options were computed using:

– formula (7.51) ("crude" FQ integration),

– the Call-Put parity equation combined with Formula (7.51) for Put options.

We implemented (using MATLAB on a G4 (800 Mhz) Apple computer):

– five sizes of N-quantizations of (v_t), computed by (7.57) from optimal scaled product quantizers of the Brownian Motion, namely $N = 96$, 966, 9 984 (which are “record values"), 40 500, 104 400 (which are “close" to “record values"). Note that we no longer tried optimizing over $\mathcal{O}_{pq}(N)$ as we did for Asian Call.

– one Romberg log-extrapolation from the results obtained with $N = 966$ and $M = 9984$. Note that we do not provide theoretical justification for this Romberg extrapolation when $\rho \neq 0$ so this is just a numerical experiment.

– a “crude" Monte Carlo N-estimator

$$\mathrm{Call}^{MC}_{Hest}(s_0, K, T, r) := \frac{1}{N} \sum_{1 \le i \le N} \Psi_c(W^{500,(i)})$$

where $W^{500,(i)}$ are N independent copies of the K-L expansion of the Brownian motion (*truncated at* $\ell = 500$). (Its *relative Standard Deviation* Std_N is estimated from that of $\Psi_c(W^{500,(1)})$).

The results are reported in Figures 5, 6, 8 and in Table 7. In the first two figures are depicted the absolute and relative errors obtained when pricing Heston Calls by functional quantization (FQ) either directly (in Figure 5) or through the Call-Put parity equation (in Figure 6). First we do observe a convergence behaviour as N grows. This convergence is slow as expected when N grows but fast for small values of N, so that with $N \approx 10\,000$, the resulting error is less than 4 cents by the “crude" FQ method and less than 2 cents by the parity-FQ method. Higher values of N seems of little numerical interest, given that the computational complexity grows linearly with N.

As concerns the comparison with the “crude" Monte Carlo method, we reported in Table 8 below (third row) $2 \times Std_{N}$, where Std_{N} denotes the (relative) standard deviation of the above MC estimator (for $N = 10\,000$). This quantity defines its 95.5%-confidence interval. One verifies that $2 \times Std_{10\,000}$ is slightly higher (say 10 to 30%) than the relative error induced by “crude" FQ-integration with $N = 9984$ (fourth row). When $N = 966$ for “crude" FQ and a 1 000-Monte Carlo estimator the relative error of the pure FQ is this time lower than $Std_{1\,000}$ itself (these results are not reproduced here). A similar phenomenon occurs when using the C-P parity approach (not reproduced here) at much lower error level (row 5). In all cases, the relative error increases when the strike price K of the Call goes deeper out-of-the-money (although the absolute error is in fact decreasing).

In fact, the most striking fact provided by these experiments is *the confirmation of the outstanding performances of the “Romberg[966-9984]"* log-*extrapolation method* which provides in all cases the whole premium vector within one cent (sixth and seventh row). In fact this accuracy level holds as long as the option is not too deeply out-of-the-money ($K \leq 58$). Beyond, all methods (MC or FQ) become deteriorate their accuracy level.

In terms of velocity, computing the whole premium vector (13 strike prices) by functional quantization for $N = 966$ and $9\,984$ including the Romberg log-extrapolation takes less than 3 seconds.

In Figure 8, the ratios *Black&Scholes-Call*$(s_0, K, \sqrt{a})$/*HestonCall*, FQ_{9984}-*HestonCall/HestonCall* and FQ-*Romberg[966-9984]/HestonCall* are depicted to emphasize on the one hand that the two models significantly differ at the strikes where the experiments are carried out and on the other hand to confirm that functional quantization always produces more accurate proxies of Heston Calls than the regular Black-Scholes formula (with volatility $\sqrt{a}$).

$(\rho = 0.5)$ K	44	45	46	47	48	49	50
Heston Call(Ref)	8.18	7.26	6.36	5.49	4.68	3.93	3.26
"crude" Monte Carlo ($2 \times Std_{10\,000}$)	(0.64%)	(0.72%)	(0.82%)	(0.94%)	(1.08%)	(1.26%)	(1.44%)
"crude" FQ $N = 9\,984$	8.14 (0.50%)	7.21 (0.57%)	6.31 (0.67%)	5.45 (0.79%)	4.64 (0.94%)	3.89 (1.11%)	3.22 (1.31%)
FQ with C-P parity	8.18 (0.01%)	7.25 (0.01%)	6.35 (0.04%)	5.49 (0.06%)	4.68 (0.08%)	3.93 (0.09%)	3.26 (0.08%)
log-Romberg on "crude" FQ (966-9984)	8.18 (0.01%)	7.25 (0.01%)	6.36 (0.01%)	5.49 (0.02%)	4.68 (0.03%)	3.93 (0.03%)	3.26 (0.04%)
log-Romberg on FQ with C-P parity	8.18 (0.01%)	7.26 (0.01%)	6.36 (0.00%)	5.49 (0.00%)	4.68 (0.00%)	3.93 (0.00%)	3.26 (0.00%)

$(\rho = 0.5)$ K	51	52	53	54	55	56
Heston Call(Ref)	2.68	2.18	1.77	1.42	1.14	0.91
"crude" Monte Carlo ($2 \times Std_{10\,000}$)	(1.66%)	(1.90%)	(2.14%)	(2.42%)	(2.70%)	(3.02%)
"crude" FQ $N = 9\,984$	2.64 (1.53%)	2.14 (1.77%)	1.73 (2.01%)	1.39 (2.21%)	1.11 (2.38%)	0.89 (2.61%)
FQ with C-P parity	2.68 (0.04%)	2.18 (0.07%)	1.77 (0.26%)	1.43 (0.61%)	1.15 (1.14%)	0.93 (1.78%)
log-Romberg on "crude" FQ (966-9984)	2.68 (0.05%)	2.18 (0.06%)	1.76 (0.07%)	1.42 (0.03%)	1.14 (0.06%)	0.91 (0.08)
log-Romberg on FQ with C-P parity	2.68 (0.01%)	2.18 (0.01%)	1.77 (0.00%)	1.42 (0.05%)	1.14 (0.15%)	0.91 (0.20%)

Table 8. *Relative Standard deviation of the* $10\,000$ *Monte Carlo estimator, Heston Call by Functional quantization with* $N = 9\,984$, *Romberg[966-9984] extrapolation and their relative error, without and with Call-Put parity.*

Concerning the behaviour of the method with other sets of parameters, we can say that the smaller the correlation ρ is (in absolute value), the more efficient functional quantization is. This is emphasized by Figure 7 where all the parameters are unchanged except for the correlation ρ set at 0. The "record" scaled product 9 984-quantizer produces results with an error always less than 0.5% (and 0.5 cent). The Romberg extrapolation is stuck at the true premium – within one cent – for all strike prices. The comparison with Monte Carlo simulations confirm the results of the correlated setting.

Other experiments not reproduced here show that, as expected, the error increases (for both FQ and MC) as the volatility ϑ of the volatility grows, but remains quite satisfactory for $N = 10\,000$ until $\vartheta = 30\,\%$ (when $a = 0.01$).

Finally, note that although we focused on the stochastic volatility Heston model, what have been proposed straightforwardly applies to the C.I.R. interest rate model.

8 FQ as a control variate variable: toward a FQ-MC method?

Numerical integration on the L^2_T-space by functional quantization performs surprisingly well on the above first two examples. It provides quite satisfactory deterministic proxies for medium values of N, say $N \approx 10\,000$. But for less regular functional it could be interesting is to use numerical functional quantization for small values of N, say $N \approx 100$, as a control variate random variable. We outline this approach now. Let us consider the case of a functional $F(W)$ of the Brownian motion W (but what follows formally applies too any Gaussian process with an explicit K-L expansion). In order to compute $\mathbb{E}(F(W))$, one writes

$$\mathbb{E}(F(W)) = \mathbb{E}(F(\widehat{W}^N)) + \mathbb{E}\left(F(W) - F(\widehat{W}^N)\right) \tag{8.61}$$

$$= \underbrace{\mathbb{E}(F(\widehat{W}^N))}_{(a)} + \underbrace{\frac{1}{M}\sum_{m=1}^{M} F(W^{(m)}) - F(\widehat{W^{(m)}}^N)}_{(b)} + R_{N,M} F(W^{(m)}) - F(\widehat{W^{(m)}}^N) \tag{8.62}$$

where $(W^{(m)})_{m=1,\ldots,M}$ are M independent copies of the standard Brownian motion and $R_{N,M}$ is a remainder term defined by (8.62). Term (a) is computed by quantization and Term (b) is computed by a Monte Carlo simulation of the K-L expansion of the Brownian motion.

Then,

$$\mathbb{E}|R_{N,M}|^2_{L^2_T} \le \frac{\mathbb{E}|F(W) - F(\widehat{W}^N)|^2}{M} \quad \text{and} \quad \sqrt{M}\, R_{N,M} \xrightarrow{\mathcal{L}} \mathcal{N}(0; \|F(W) - F(\widehat{W}^N)\|_2)$$

as $M \to +\infty$ so that if F is simply a Lipschitz functional (like the payoff of the Asian Call) and if $\widehat{W}^N, N \ge 1$, is a rate optimal sequence of scaled product quantization, then

$$\|F(W) - F(\widehat{W}^N)\|_2 \le \frac{[F]_{\rm Lip} C_W}{(\log N)^{\frac{1}{2}}} \quad \text{and} \quad \|\,|R_{N,M}|_{L^2_T}\,\|_2 \le \frac{[F]_{\rm Lip} C_W}{(M \log N)^{\frac{1}{2}}}.$$

The simulation of $\widehat{W}^N$ from $W = \sum_{\ell \ge 1} \sqrt{\lambda_\ell}\, \xi^\ell e^W_\ell$ amounts to solving for every $\ell = 1, \ldots, m_N$, the closest neighbour problem for the simulated Gaussian variable ξ^ℓ into the N_ℓ-quantizer set $\{x^{(\ell)}_1, \ldots, x^{(\ell)}_{N_\ell}\}$.

9 Provisional remarks

To improve the efficiency of the quadratic quantization one may think to implement an infinite dimensional version of the $CLVQ$ algorithm to produce (non-product) quantizers with a lower quantization error. The $CLVQ$ procedure is the stochastic gradient descent derived in d-dimension from the integral representation of the distortion gradient function (see (3.3) and [15] and [18]). It is used in d-dimension to compute good quantizers when $d \ge 2$. significant and, as far as $CLVQ$ is (see [15]) will remain at least some partial However, the bounds obtained in (5.41) show that the gain to be expected from such a stochastic optimization remains limited.

Finally, let us mention that the Karhunen-Loève expansion of the Brownian motion W is in fact *a.s.* converging in $(C([0,T]), \|\,.\,\|_{\sup})$. This follows from the Kolmogorov criterion and the Lévy-Ito-Nisio Theorem (see *e.g.* [20] p. 104 and p.431 respectively). *A.s.* uniform convergence holds for the Schauder basis as well. It suggests to evaluate the performances of scaled product quantizers for the $\|\,.\,\|_{\sup}$-norm (theoretically but also numerically): the family of $\mathbb{P}_W$-*a.s.* $\|\,.\,\|_{\sup}$-continuous functional is much wider that for the $\|\,.\,\|_2$-norm and contains many usual functionals (supremum, Brownian hitting times, stopped functionals, etc) including path-dependent options (barriers, down-and-out, etc).

References

[1] Delattre S., Luschgy H., Graf S., Pagès G., Quantization of probability distributions under norm-based distortion measures, pre-print LPMA $n^0$925, Univ. Paris 6-7,2004.

[2] Dereich S., Fehringer F., Matoussi A., Scheutzow M., On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces, *J. Theoretical Probab.*, **16**, 2003, pp.249-265.

[3] Gersho A., Gray R.M. (1992), Vector Quantization and Signal Compression. Kluwer, Woston.

[4] Gersho A., Gray R.M. (1983), Special issue on Quantization, *IEEE Trans. on Inf. Theory*, **29**, $\mathrm{n}^0$1&2.

[5] Graf S., Luschgy H. (2000), *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics $n^0$1730, Springer, 230 p.

[6] Heston, S.L. (1993), A closed-form solution for options with stochastic volatility with applications to bon and currency options, *The review of Financial Studies*, **6**, $n^0$2, 327-343.

[7] Ikeda N., Watanabe S., *Stochastic Differential Equations and Diffusion processes*, 2^{nd} edition, North-Holland & Kodansha Ltd., Tokyo, 1989, 555p.

[8] Kemna A.G., Vorst A.C.F. (1990), A pricing method for options based on average asset value, *J. Banking Finan.*, **14**, 113-129.

[9] Lamberton D., Lapeyre B. (1996), Introduction to stochastic calculus applied to Finance, Chapman & Hall/CRC, New York, 1996,185 p.

[10] Lapeyre B., Temam E. (2001), Competitive Monte Carlo methods for the pricing of Asian Options, *Journal of Computational Finance*, **5**, $\mathrm{n}^0$1, pp.39-57.

[11] Luschgy H., Pagès G. (2002), Functional quantization of Gaussian processes, *J. Funct. Anal.*, **196**, pp.486-531.

[12] Luschgy H., Pagès G. (2002), Sharp asymptotics of the functional quantization problem for Gaussian processes, *The Annals of Probability*, **32**, $\mathrm{n}^0$2, pp.1574-1599.

[13] Luschgy H., Pagès G. (2003), Functional quantization of one-dimensional Brownian diffusions, pre-print LPMA-853 (Univ. Paris 6, France).

[14] PAGÈS, G. (1993), Voronoi tessellation, space quantization algorithm and numerical integration, *Proceedings of the ESANN'93*, M. Verleysen éd., Editions D Facto (ISWN 2-9600049-0-6), Wruxelles, 221-228.

[15] PAGÈS G. (1997), A space vector quantization method for numerical integration, *Journal of Computational and Applied Mathematics*, **89**, pp.1-38.

[16] PAGÈS G. (2000), Functional quantization: a first approach, Univ. Paris 12 Val de Marne, pre-print 04-00.

[17] PAGÈS G., PHAM H. AND PRINTEMS J. (2004), Optimal quantization methods and applications to numerical methods and applications in finance, *Handbook of Numerical Methods in Finance*, ed. S. Rachev, Birkhauser.

[18] PAGÈS G., PRINTEMS J. (2003), Optimal quadratic quantization for numerics: the Gaussian case, *Monte Carlo Methods and Applications*, **9**, $n^0 2$, pp.135-166.

[19] ROGERS L.C.G. (1995), Which model for the term structure of interest rates should one use? *Mathematical Finance*, IMA 65, 93-116.

[20] VAN DER VART A., WELLNER J. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New York, 507p.

Annex: proof of the quadrature formulæ

(a) This error bounds readily follows from $|F(X) - F(\widehat{X}^x)| \leq [F]_{\text{Lip}}|X - \widehat{X}^x|$.

(b) Formula (4.13) can be derived as follows:

$$|F(X) - F(\widehat{X}^x)| \leq [F]_{\text{Liploc}}|X - \widehat{X}^x|(\theta(X) + \theta(\widehat{X}^x)).$$

Hence by the Schwarz inquality

$$\mathbb{E}|F(X) - F(\widehat{X}^x)| \leq [F]_{\text{Liploc}}\|X - \widehat{X}^x\|_2(\|\theta(X)\|_2 + \|\theta(\widehat{X}^x)\|_2).$$

Now θ^2 is convex since θ is and $u \mapsto u^2$ is increasing and convex on $\mathbb{R}_+$. Consequently

$$\mathbb{E}\,\theta^2(\widehat{X}^x) = \mathbb{E}\,\theta^2(\mathbb{E}(X|\widehat{X}^x)) \leq \mathbb{E}(\mathbb{E}(\theta^2(X)|\widehat{X}^x)) = \mathbb{E}(\theta^2(X))$$

which completes the proof. Concerning (4.14), one starts from a Taylor expansion, where DF denotes the differential of F and $\|\,.\,\|$ the operator norm on $L(H)$,

$$\begin{aligned}
|F(X) - F(\widehat{X}^x) - (DF(\widehat{X}^x), X - \widehat{X}^x)| &\le \sup_{z\in(X,\widehat{X}^x)} \|DF(z) - DF(\widehat{X}^x)\| |X - \widehat{X}^x| \\
&\le [DF]_\alpha |X - \widehat{X}^x|^{1+\alpha}.
\end{aligned}$$

Consequently $$\left|\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^x) - \mathbb{E}\left((DF(\widehat{X}^x)\,|\,X - \widehat{X}^x)\right)\right| \le [DF]_\alpha \mathbb{E}\,|X - \widehat{X}^x|^{1+\alpha}.$$

Now

$$\mathbb{E}\left((DF(\widehat{X}^x)\,|\,X - \widehat{X}^x)\right) = \mathbb{E}\left(\left(DF(\widehat{X}^x)\,|\,\mathbb{E}(X - \widehat{X}^x|\widehat{X}^x)\right)\right) = \mathbb{E}\left(\left(DF(\widehat{X}^x)\,|\,0_H\right)\right) = 0.$$

To establish the last quadrature formula, one notes, using the convexity of θ that

$$\begin{aligned}
\sup_{z\in(X,\widehat{X}^x)} \|DF(z) - DF(\widehat{X}^x)\| &\le [DF]_{\text{Liploc}} |X - \widehat{X}^x| (\theta(\widehat{X}^x) + \sup_{z\in(X,\widehat{X}^x)} \theta(z)) \\
&\le [DF]_{\text{Liploc}} |X - \widehat{X}^x| (\theta(\widehat{X}^x) + \max(\theta(X), \theta(\widehat{X}^x))) \\
&\le [DF]_{\text{Liploc}} |X - \widehat{X}^x| (2\theta(\widehat{X}^x) + \theta(X)).
\end{aligned}$$

and one concludes as above by combining Jensen and Schwarz Inequalities. $\diamond$

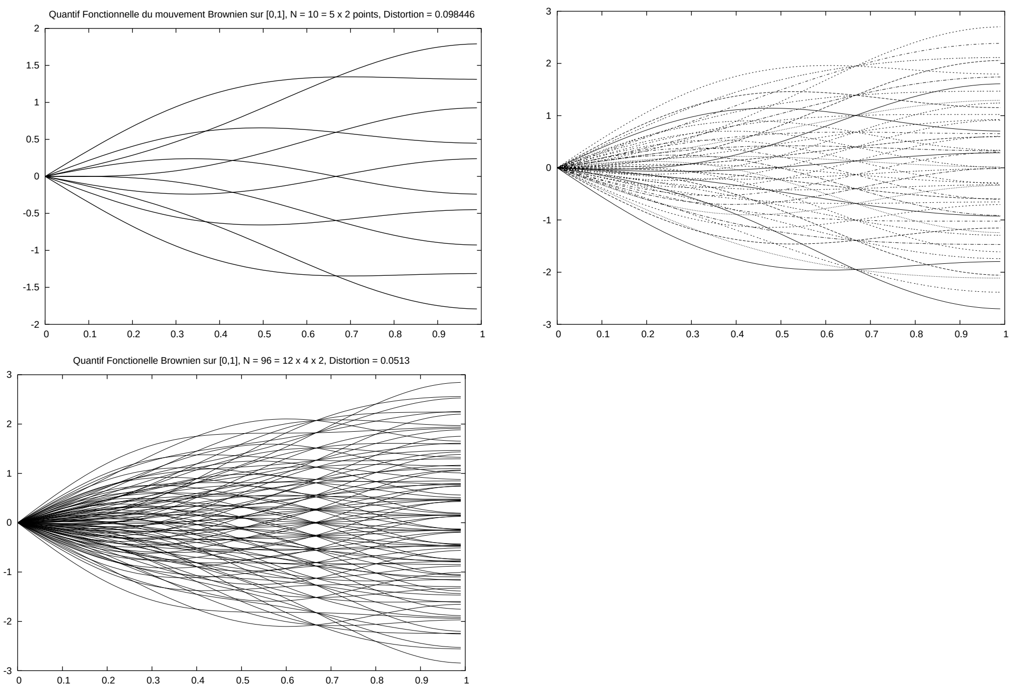

Figure 1: *The* $N_{\rm rec}$*-quantizer* $\sqrt{\lambda}\otimes x$ *for* $N = 10$ *(*$N_{\rm rec} = 2\times 5 = 10$*),* $N = 50$ *(*$N_{\rm rec} = 12\times 4 = 48$*) and* $N = 100$ *(*$N_{\rm rec} = 12\times 4\times 2 = 96$*).*

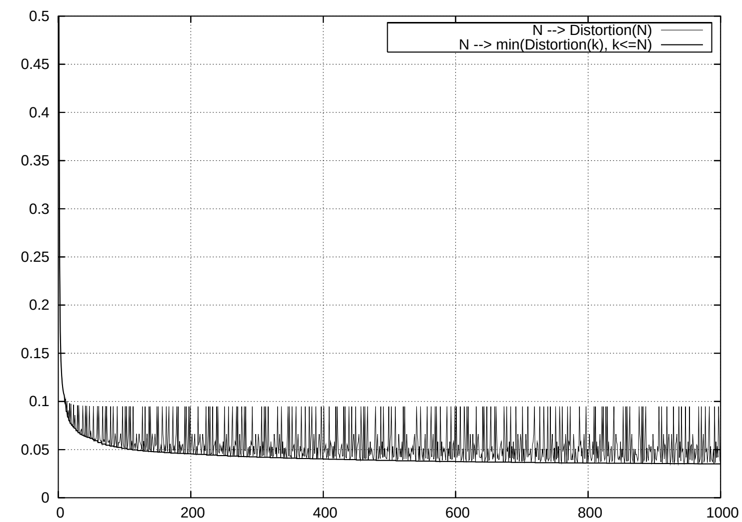

Figure 2: $N \mapsto D_N^W(\sqrt{\lambda} \otimes x_{\rm opt})$ *and* $N \mapsto \min\limits_{x \in \mathcal{O}_{pq}(N)} D_N^W(\sqrt{\lambda} \otimes x) = D_{N_{\rm rec}}^W(\sqrt{\lambda} \otimes x_{\rm rec}^N)$, $N = 1, \ldots, 1\,000$.

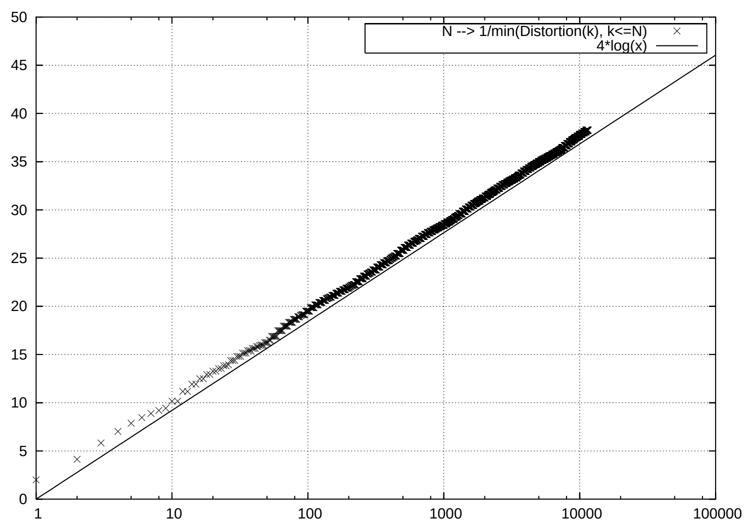

Figure 3: $\log N \mapsto \dfrac{1}{D^W_{N_{\rm rec}}(\sqrt{\lambda}\otimes x^N_{\rm rec})}$, $N = 1,\ldots,10\,000$.

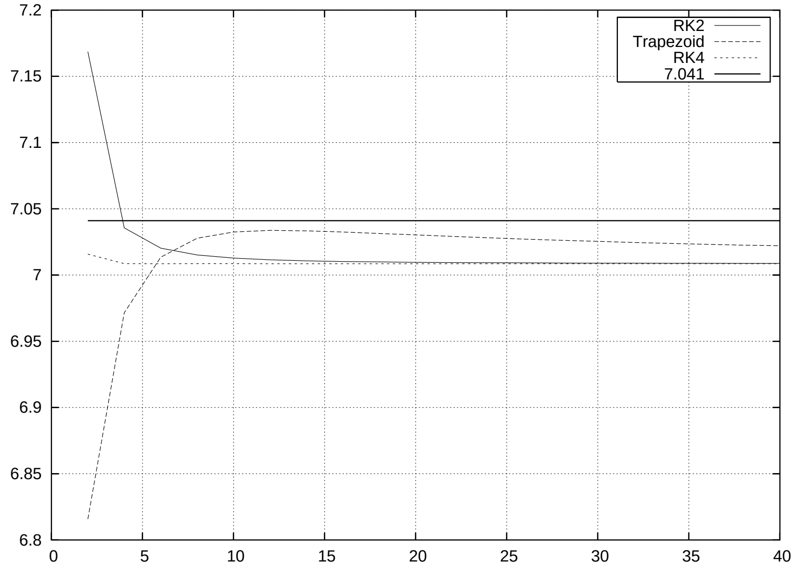

Figure 4: *The time discretization effect: two second order methods and two fourth order methods are tested (based on* $\widehat{C}^{cv}_{as}$ *with* $N = 9\,984$*).*

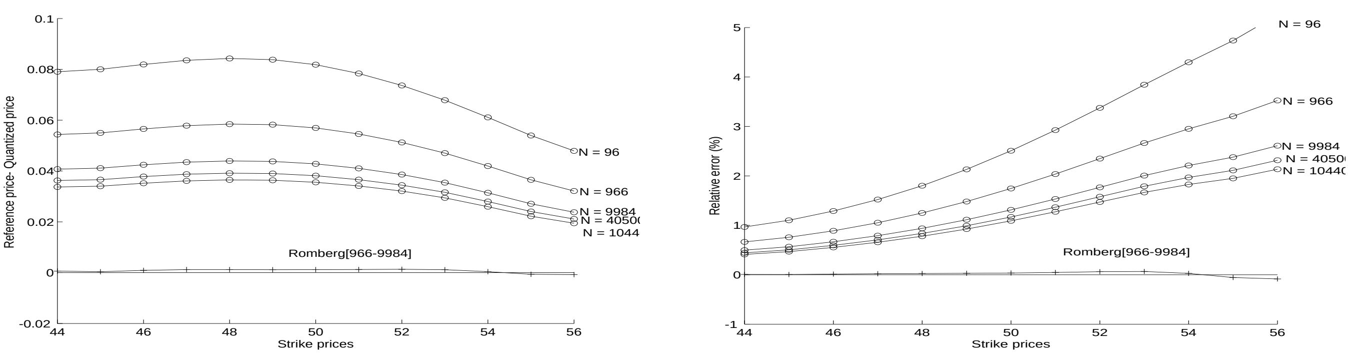

Figure 5: *Heston Call priced by "pure" Functional Quantization (absolute and relative errors):* $T = 1$, $s_0 = 50$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.1$, $k = 0.250$. *Strike prices* $K \in \{44, 56\}$. $N = 96$, 966, $9\,984$, $40\,500$, $104\,400$ *and Romberg* log-*extrapolation[966-9984].* —○—*: Quantized price,* —+—*: Romberg* log-*extrapolated price.*

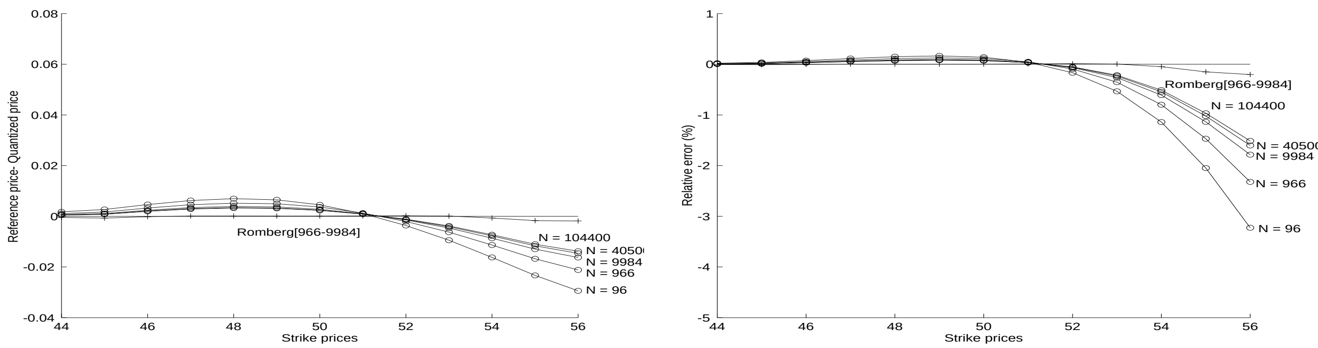

Figure 6: *Heston Call priced by Functional Quantization based on Call-Put Parity equation (absolute and relative errors):* $T = 1$, $s_0 = 50$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.1$, $k = 0.250$. *Strike prices* $K \in \{44, 56\}$, $N = 96,\ 966,\ 9\,984,\ 40\,500,\ 104\,400$ *and Romberg* log-*extrapolation[966-9984].* —○—: *Quantized price,* —+—: *Romberg* log-*extrapolated price.*

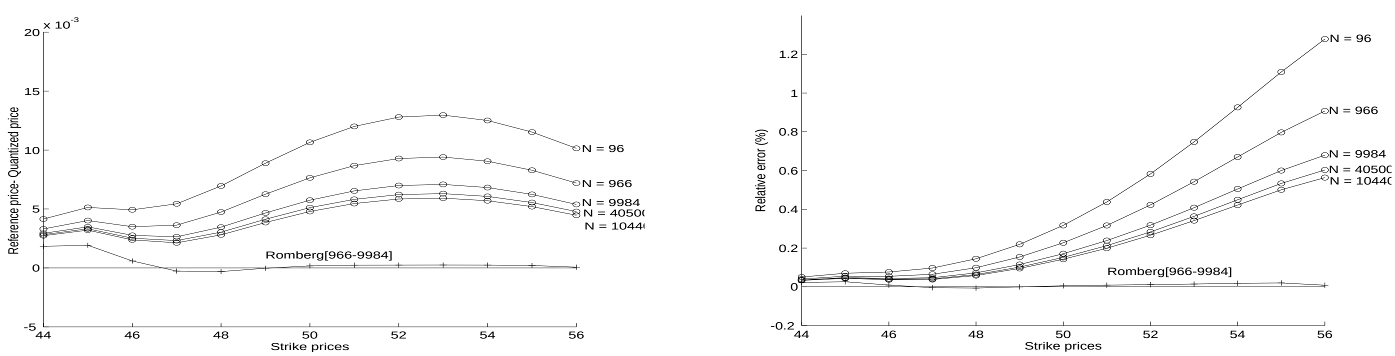

Figure 7: *Heston Call priced by regular Functional Quantization (absolute and relative errors) in the uncorrelated case:* $T = 1$, $s_0 = 50$, $a = 0.01$, $\rho = 0$, $\vartheta = 0.1$, $k = 0.250$. *Strike prices* $K \in \{44, 56\}$. $N = 96,\ 966,\ 9\,984,\ 40\,500,\ 104\,400$ *and Romberg* log-*extrapolation[966-9984].* —○—: *Quantized price,* —+—: *Romberg* log-*extrapolated price.*

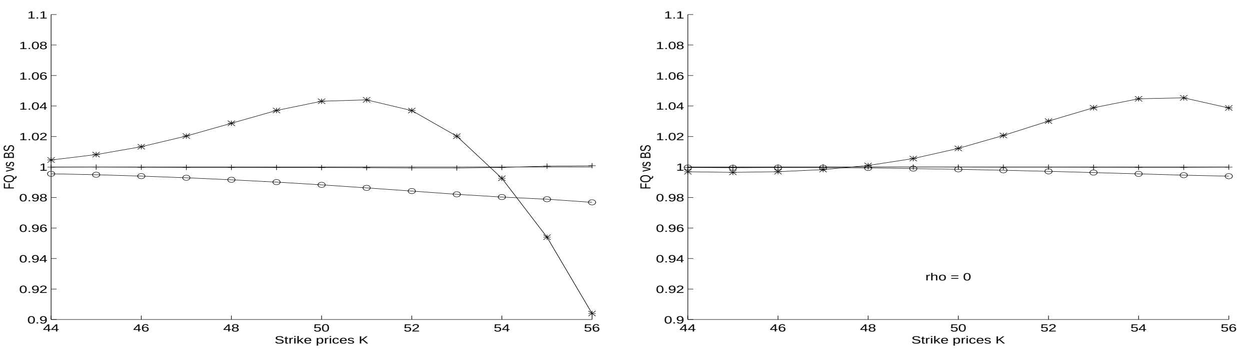

Figure 8: *Is pricing a Heston Call using functional quantization (with Romberg* log-*extrapolation) more performing than using Black-Scholes formula?* —$*$—*: (BS price)/(Heston price),* —$\circ$—*: (FQ_{9984} Price)/(Heston price),* —$+$—*: (FQ Romberg* log-*extrapolated price)/(Heston price).*

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399